

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 14:47:39 ; Search time 53.42 Seconds
(without alignments)
95.646 Million cell updates/sec

Title: US-09-833-017-2

Sequence: 1 MKRTSLKNDKFKITDEP.....GSLSTFFELNRSFTQALGK 46

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

A.GeneSeq_032802:*

- 1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	58	25.6	173	AAW28307	Staphylococcus aur
2	56	24.7	939	AAU35204	Enterococcus faeca
3	55.5	24.4	553	AAU70176	Sequence of Newcas
4	55.5	24.4	553	AAU88858	Newcastle Disease
5	55.5	24.4	553	AAU9141	Newcastle Disease
6	55.5	24.4	553	AAU06828	Newcastle Disease
7	55.5	24.4	553	AAU10691	Newcastle Disease
8	55.5	24.4	553	AAU21983	Seq ID No: 14 of U
9	55.5	24.4	553	AAU36039	Protein encoded by
10	55.5	24.4	553	AAU51231	Newcastle disease
11	55.5	24.4	553	AAU58183	NDV fusion (F) pro

12	55	24.2	583	22	AAU82132	S. epidermidis ope
13	54.5	24.0	508	21	AAU44399	Arabidopsis thalia
14	54.5	24.0	515	21	AAU44398	Arabidopsis thalia
15	54	23.8	96	14	AAU35757	Monellin protein.
16	54	23.8	97	14	AAU34192	Monellin protein.
17	54	23.8	2353	17	AAU93933	Haemophilus adhe
18	54	23.8	2411	21	AAU23860	Haemophilus influe
19	53.5	23.6	553	12	AAU10065	Newcastle disease
20	53	23.3	96	21	AAU57201	Arabidopsis thalia
21	53	23.3	513	15	AAU54214	L. lactis branched
22	52	22.9	148	22	AAU28821	Novel human diago
23	52	22.9	365	22	AAU78620	Human protein SEQ
24	52	22.9	381	22	AAU79604	Human protein SEQ
25	52	22.9	381	22	AAU79605	Human protein SEQ
26	52	22.9	382	22	AAU78621	Human protein SEQ
27	52	22.9	516	21	AAU52007	M. jannaschii MJ08
28	52	22.9	516	21	AAU51636	M. jannaschii MJ08
29	51.5	22.7	177	22	AAU17731	Novel human diago
30	51.5	22.7	257	22	AAU98329	Escherichia coli p
31	51.5	22.7	306	21	AAU5918	E. coli proliferat
32	51.5	22.7	306	22	AAU98991	E. coli growth and
33	51.5	22.7	591	22	AAU10345	Novel human diago
34	51.5	22.7	2737	22	AAU18288	Novel human diago
35	51.5	22.7	3076	22	AAU07038	Novel human diago
36	51	22.5	220	22	AAU37294	Staphylococcus aur
37	51	22.5	292	22	AAU74939	Human colon cancer
38	51	22.5	313	18	AAU21821	Mutant uracil DNA
39	51	22.5	313	18	AAU21822	Mutant uracil DNA
40	51	22.5	313	18	AAU21824	Mutant uracil DNA
41	51	22.5	313	18	AAU21825	Mutant uracil DNA
42	51	22.5	313	18	AAU21814	Uracil DNA glycosy
43	51	22.5	313	18	AAU21815	Cytosine DNA glyco
44	51	22.5	313	18	AAU21816	Thymine DNA glycos
45	51	22.5	313	18	AAU21819	Mutant uracil DNA

ALIGNMENTS

RESULT 1	AAW28307	standard; Protein: 173 AA.
ID	AAW28307;	
AC	AAW28307;	
XX		
DT	01-SEP-1998 (first entry)	
XX		
DE	Staphylococcus aureus protein of unknown function.	
XX		
KW	Staphylococcus aureus protein; ribozyme; antisense sequence; control;	
KW	Staphylococcal gene; regulatory element; bacterial gene expression;	
KW	vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;	
KW	toxic shock syndrome.	
XX		
OS	Staphylococcus aureus.	
XX		
FT	Key	location/qualifiers
FT	Misc-difference 123	/note= "X is not defined in the specification"
FT	Misc-difference 132	/note= "X is not defined in the specification"
FT	Misc-difference 132	/note= "X is not defined in the specification"
XX		
PN	MO9730070-A1.	
XX		
PD	21-AUG-1997.	
XX		
PF	19-FEB-1997; 97WO-US02318.	
XX		
PR	20-FEB-1996; 96US-0011888.	
XX		
PA	(SMK) SMITHKLINE BEECHAM CORP.	
XX		
PI	Black MT, Burnham MK, Hodgson JB, Knowles DJC, Nicholas RO;	

```

PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;
XX
XX WPI: 1997-424969/39.
DR N-PSDB; AAT84204.
XX
PT Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used
PT to isolate antimicrobial compounds, and in vaccines against S.
PT aureus infection
XX
XX Claim 6; Page 597-598; 989pp; English.
XX
CC The present sequence represents a Staphylococcus aureus protein of
CC unknown function. The DNA sequence was isolated from a library of
CC clones of S. aureus WCUH 29 in Escherichia coli. The DNA sequence can
CC be used in the construction of ribozymes and antisense sequences to
CC control the expression of Staphylococcal genes. The DNA sequence is
CC also useful as a source of regulatory elements for the control of
CC bacterial gene expression. The present protein may be used to produce
CC vaccines to enable a host to produce specific antibodies with
CC antibacterial action. These vaccines and antibodies would protect
CC a host against invasion by S. aureus, and conditions relating to
CC Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled
CC skin syndrome, and toxic shock syndrome.
XX
XX Sequence 173 AA;
SO

```

Query Match 25.6%; Score 58; DB 18; Length 173;
 Best Local Similarity 34.3%; Pred. No. 3.3;
 Matches 12; Conservative 8; Mismatches 15; Indels 0; Gaps 0;

```

OY 9 NDFKEIKTDELEIIGSGSLTFEFLFRNRSFTQA 43
DB 86 ndfpelfdcltlnvisankdltkwtfqynkellsa 120

```

RESULT 2
 AAU35204
 ID AAU35204 standard; Protein: 939 AA.
 AC AAU35204;
 DT 13-FEB-2002 (first entry)
 DE Enterococcus faecalis cellular proliferation protein #491.
 XX Antisense; prokaryotic cellular proliferation protein;
 KW antibiotic; antibacterial; drug design.
 OS Enterococcus faecalis.
 XX WO200170955-A2.
 PN
 PD 27-SEP-2001.
 XX
 PF 21-MAR-2001; 2001WO-US09180.
 XX
 PR 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Haselbeck R, Ohlsen KL, Zyskind JM, Wall D, Trawick JD, Carr CJ;
 PI Yamamoto RT, Xu HH;
 XX WPI: 2001-611495/70.
 DR N-PSDB; AAS53063.
 XX

```

PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 10797; 51pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 939 AA;
SO

```

Query Match 24.7%; Score 56; DB 22; Length 939;
 Best Local Similarity 33.3%; Pred. No. 45;
 Matches 11; Conservative 7; Mismatches 15; Indels 0; Gaps 0;

```

OY 5 LSKNDFKEIKTDELEIIGSGSLTFEFLFRN 37
DB 330 idmdtprfelpdqgqellngseknfhlyen 362

```

RESULT 3
 AAP70176
 ID AAP70176 standard; Protein: 553 AA.
 AC AAP70176;
 DT 03-APR-1991 (first entry)
 DE Sequence of Newcastle Disease Virus (NDV) F polypeptide.
 XX Epitope; probe; diagnosis.
 KW
 OS Newcastle Disease Virus.
 XX EP227414-A.
 PN
 PD 01-JUL-1987.
 XX
 PF 16-DEC-1986; 86EP-0309804.
 XX
 PR 15-JUL-1986; 86US-0885765.
 PR 18-DEC-1985; 85GB-0031147.
 PR 14-APR-1986; 86GB-0009037.
 XX
 PA (NATR) NATIONAL RES DEV CORP.
 XX
 PI Bingham RW, Chambers P, Emerson PT, Millar NS;
 XX WPI: 1987-179630/26.
 DR N-PSDB; AAN70261.
 XX
 PT Newcastle disease virus gene clones - comprise polynucleotide(s)
 PT encoding the HN and/or F protein of Newcastle disease virus RNA
 XX
 PS Example; pages 11-16; 22pp; English.
 XX

[illegible]

XX	Newcastle disease virus.
OS XX	
PX	M09640880-A1.
PN XX	
PD XX	19-DEC-1996.
PF XX	
PR XX	04-JUN-1996; 96MO-US11187.
PS XX	
PT XX	07-JUN-1995; 95US-0484790.
PI XX	(SYTR) SYNTRO CORP.
PA XX	Cochran MD, Junker DE, Singer PA;
PP XX	WPI: 1997-087060/08.
DR XX	N-PSTD; AAT48510.
DT XX	
DE XX	New recombinant fowlpox virus - contg. a foreign DNA sequence inserted into the fowlpox virus genome, used for the prodn. of vaccines.
DV XX	
SQ XX	Disclosure: Page 109-110; 134pp; English.
CC CC	Newcastle disease virus (NDV) haemagglutinin (HN) (AAU10690) and fusion (F) protein (AAY10691) are expressed by novel recombinant cowpox virus (CPV). The genes (see also AAU48510) for HN and F can be inserted into homology vector 443-88.8 (see also AAU48511) at the unique SfiI site, yielding homology vector 502-26.22 (see also AAU48502-05), which can be used to insert the NDV HN and F genes into fowlpox virus. The recombinant virus is used to deliver the vaccine antigens to poultry.
CC CC	
CS XX	Sequence 553 AA:
SC XX	
YQ	Query Match 24.4%; Score 55.5; DB 18; Length 553; Best Local Similarity 33.3%; Pred.No. 29; Matches 16; Conservative 7; Mismatches 20; Indels 5; Gaps 4 TSLKDNDF-----KEIKTDELIIIGSGSLSLTFRLFNRSFTQLAK 46 ::: : :: : : Db 433 tlrslgefdvlyqknisdsgvltnldisteigvnmslnalnkk 480
RESULT	8
ID	AAY21983 standard; Protein: 553 AA. AC AAY21983; AD 07-SEP-1999 (first entry) AE Seq ID No: 14 of US5923558. AF AG Fowlpox virus; FPV; recombinant; Vaccine; Immunisation; Chicken; NDV AH Newcastle disease virus; NDV; Fowlpox; Infectious Laryngotracheitis. AI Fowlpox virus. AJ U55923558-A. AK PN PD 20-JUL-1999. AL PF 07-JUN-1995; 95US-0484575. AM PR 07-JUN-1995; 95US-0484575. AN PR 26-FEB-1993; 93US-0024156. AO PR 28-FEB-1994; 94WO-US025252. AP XX (SVTR) SYNTRO CORP. AQ PA Cochran MD, Junker DE, AR PI

[illegible]


```
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0150566.
PR 25-AUG-1999; 99US-0150884.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 24.0%; Score 54.5; DB 21; Length 508;
Best Local Similarity 39.6%; Pred. No. 36;
Matches 21; Conservative 8; Mismatches 11; Indels 13; Gaps 5;

```
Oy 1 MKKTLKND-FKE-----IKTDELEII-GSGSGLS---TF---FRLFRSF 40
Db 1 mekrtlkndsfvgeyprvtetglsilvgasgdlaaktkkfaiInlfnfgf 53
```

```
RESULT 14
AAG44398
ID AAG44398 standard; Protein; 515 AA.
XX
AC AAG44398;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 55611.
XX
KW Protein identification; signal transduction pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138547.
PR 10-JUN-1999; 99US-0138840.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
```


PR	18-JUN-1999,	990S-0139459
PR	18-JUN-1999,	990S-0139460
PR	18-JUN-1999,	990S-0139461
PR	18-JUN-1999,	990S-0139462
PR	18-JUN-1999,	990S-0139463
PR	18-JUN-1999,	990S-0139750
PR	21-JUN-1999,	990S-0139763
PR	22-JUN-1999,	990S-0139817
PR	23-JUN-1999,	990S-0139899
PR	23-JUN-1999,	990S-0140353
PR	24-JUN-1999,	990S-0140354
PR	28-JUN-1999,	990S-0140695
PR	29-JUN-1999,	990S-0140823
PR	30-JUN-1999,	990S-0140991
PR	01-JUL-1999,	990S-0141287
PR	01-JUL-1999,	990S-0141842
PR	14-JUL-1999,	990S-0142154
PR	15-JUL-1999,	990S-0143624
PR	16-JUL-1999,	990S-0144005
PR	16-JUL-1999,	990S-0144085
PR	19-JUL-1999,	990S-0144086
PR	19-JUL-1999,	990S-0144325
PR	19-JUL-1999,	990S-0144331
PR	19-JUL-1999,	990S-0144332
PR	19-JUL-1999,	990S-0144333
PR	19-JUL-1999,	990S-0144811
PR	20-JUL-1999,	990S-0144834
PR	20-JUL-1999,	990S-0144632
PR	21-JUL-1999,	990S-0144632
PR	21-JUL-1999,	990S-0144814
PR	21-JUL-1999,	990S-0145086
PR	22-JUL-1999,	990S-0145085
PR	22-JUL-1999,	990S-0145087
PR	22-JUL-1999,	990S-0145088
PR	23-JUL-1999,	990S-0145192
PR	23-JUL-1999,	990S-0145145
PR	23-JUL-1999,	990S-0145218
PR	23-JUL-1999,	990S-0145224
PR	26-JUL-1999,	990S-0145276
PR	27-JUL-1999,	990S-0145913
PR	27-JUL-1999,	990S-0145918
PR	28-JUL-1999,	990S-0145919
PR	02-AUG-1999,	990S-0146386
PR	02-AUG-1999,	990S-0146388
PR	02-AUG-1999,	990S-0146389
PR	03-AUG-1999,	990S-0147038
PR	04-AUG-1999,	990S-0147204
PR	05-AUG-1999,	990S-0147302
PR	11-AUG-1999,	990S-0147319
PR	12-AUG-1999,	990S-0148341
PR	13-AUG-1999,	990S-0148565
PR	13-AUG-1999,	990S-0148684
PR	16-AUG-1999,	990S-0149388
PR	17-AUG-1999,	990S-0149175
PR	18-AUG-1999,	990S-0149426
PR	20-AUG-1999,	990S-0149722
PR	20-AUG-1999,	990S-0149723

PR	20-AUG-1999;	9905-0149929
PR	22-AUG-1999;	9905-0149902
PR	23-AUG-1999;	9905-0149930
PR	25-AUG-1999;	9905-0105666
PR	26-AUG-1999;	9905-0105084
PR	27-AUG-1999;	9905-0151065
PR	27-AUG-1999;	9905-0151066
PR	27-AUG-1999;	9905-0151080
PR	30-AUG-1999;	9905-0151303
PR	31-AUG-1999;	9905-0151438
PR	01-SEP-1999;	9905-0151930
PR	07-SEP-1999;	9905-0152630
PR	10-SEP-1999;	9905-0153708
PR	13-SEP-1999;	9905-0153758
PR	15-SEP-1999;	9905-0154018
PR	16-SEP-1999;	9905-0154039
PR	20-SEP-1999;	9905-0154779
PR	22-SEP-1999;	9905-0155133
PR	23-SEP-1999;	9905-0155139
PR	24-SEP-1999;	9905-0155486
PR	28-SEP-1999;	9905-0156458
PR	29-SEP-1999;	9905-0156596
PR	04-OCT-1999;	9905-0157117
PR	05-OCT-1999;	9905-0157753
PR	06-OCT-1999;	9905-0157865
PR	07-OCT-1999;	9905-0158029
PR	08-OCT-1999;	9905-0158232
PR	12-OCT-1999;	9905-0158366
PR	13-OCT-1999;	9905-0159293
PR	13-OCT-1999;	9905-0159294
PR	13-OCT-1999;	9905-0159295
PR	14-OCT-1999;	9905-0159329
PR	14-OCT-1999;	9905-0159330
PR	14-OCT-1999;	9905-0159531
PR	14-OCT-1999;	9905-0159537
PR	18-OCT-1999;	9905-0159583
PR	18-OCT-1999;	9905-0159584
PR	21-OCT-1999;	9905-0160744
PR	21-OCT-1999;	9905-0160767
PR	21-OCT-1999;	9905-0160768
PR	21-OCT-1999;	9905-0160770
PR	21-OCT-1999;	9905-0160815
PR	22-OCT-1999;	9905-0160980
PR	22-OCT-1999;	9905-0160981
PR	22-OCT-1999;	9905-0160989
PR	25-OCT-1999;	9905-0161404
PR	25-OCT-1999;	9905-0161405
PR	25-OCT-1999;	9905-0161406
PR	26-OCT-1999;	9905-0161359
PR	26-OCT-1999;	9905-0161360
PR	28-OCT-1999;	9905-0161361
PR	28-OCT-1999;	9905-0161920
PR	28-OCT-1999;	9905-0161993
PR	29-OCT-1999;	9905-0162142

```

Query Match          24.0%;      Score 54.5; DB 21; Length 515;
Best Local Similarity 39.6%;
Matches 21; Conservative 8; Mismatches 11; Indels 13; Gaps 5

QY 1 MKKTLISKMD-EKE-----IKTDELEIIT-GSGGSLG---TF---RRLFNRSF 40
   | : | : | | | | : | : | : | : | : | : | : | : | : | : | : |
Db 8 mekrtstlksndsfvgeynpyvtelcgsisllylgasgdiakktkfpaflnlhgaf 60

RESULT 15
AAR35757
ID AAR35757 standard; Protein; 96 AA.
XX
AC AAR35757;
XX
DT 29-JUL-1993 (first entry)

```

```

XX Monellin protein.
DE
XX Sweet; heat; resistance; protease.
KW
XX Synthetic.
OS
XX Key Location/Qualifiers
FH Misc-difference 51
FT /label= GLY, SER, GLU
XX
XX JP05070494-A.
PN
XX
XX 23-MAR-1993.
PD
XX
XX 30-MAY-1991; 91JP-0155713.
PF
XX 25-JUL-1990; 90JP-0196983.
PR
XX (KIRI ) KIRIN BREWERY KK.
PA
XX
XX WPI; 1993-131302/16.
DR
XX
XX Single-stranded monellin protein - has high heat resistance in
PT cow PH range and high protease resistance
PT
XX Claim 1; Page 2; 29pp; Japanese.
PS
XX Monellin, is a sweet protein, having high thermal resistance at low
CC PH's and also a high protease resistance.
CC
XX
SQ Sequence 96 AA;

```

```

Query Match 23.8%; Score 54; DB 14; Length 96;
Best local Similarity 41.4%; Pred. No. 6;
Matches 12; Conservative 6; Mismatches 11; Indels 0; Gaps 0;
OY 1 MKKTLSLKNDPKFKIKDELEIITGGSGSL 29
    ||||: :|:|:| | : : : |
Db 42 mkktiyeenxfreikgyeyqlvyvasdki 70

```

Search completed: July 30, 2002, 15:14:11
Job time: 1592 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:06:04 ; Search time 23.35 seconds

(without alignments)
48.119 Million cell updates/sec

Title: US-09-833-017-2

Perfect score: 227 1 MKKTSLKNDKFEKITDELE.....GSLSTFFLENRSTQALGK 46

Sequence:

BIOSUM62
Gapop 10.0 , Gapext 0.5

Scoring table:

Searched: 231628 seqs, 24425594 residues

231628

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: Issued_Patents_AA:*
2: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	55.5	24.4	553	2	US-08-663-566A-13
2	55.5	24.4	553	2	US-08-484-575A-14
3	55.5	24.4	553	2	US-08-023-610-13
4	55.5	24.4	553	2	US-08-288-065A-13
5	55.5	24.4	553	2	US-08-362-240A-13
6	55.5	24.4	553	3	US-08-477-459-14
7	55.5	24.4	553	3	US-08-479-869-14
8	55.5	24.4	553	4	US-08-486-414-14
9	55.5	24.4	553	4	US-08-804-372A-11
10	55.5	24.4	553	5	PCT-US94-01826A-14
11	55.5	24.4	553	5	PCT-US94-02252A-14
12	55.5	24.4	553	5	PCT-US95-10245-13
13	55.5	24.4	553	6	5310678-1
14	55.5	24.4	837	4	US-09-564-805-228
15	54	23.8	97	4	US-09-242-690A-6
16	54	23.8	1912	1	US-08-409-995-4
17	54	23.8	1912	3	US-08-685-467-4
18	54	23.8	2353	4	US-09-377-155-33
19	54	23.8	2353	4	US-08-913-942-4
20	54	23.8	2353	4	US-09-669-974-33
21	54	23.8	2354	4	US-09-268-347-47
22	54	23.8	2411	4	US-09-268-347-36
23	53	23.3	513	1	US-08-403-866-1
24	51.5	22.7	244	4	US-08-919-573-2
25	51.5	22.7	244	4	US-08-919-573-2
26	49.5	21.8	436	3	US-08-486-099-94
27	49.5	21.8	436	3	US-08-360-107A-104

28	49.5	21.8	436	3	US-08-484-223B-94	Sequence 94, Appl
29	49.5	21.8	436	3	US-08-919-597-94	Sequence 94, Appl
30	49.5	21.8	436	3	US-08-475-668A-94	Sequence 94, Appl
31	49.5	21.8	436	3	US-08-485-551A-94	Sequence 94, Appl
32	49.5	21.8	436	3	US-08-471-913A-94	Sequence 94, Appl
33	49.5	21.8	436	4	US-08-485-264A-94	Sequence 94, Appl
34	49.5	21.8	436	4	US-08-474-349A-94	Sequence 94, Appl
35	48.5	21.4	202	4	US-08-858-207A-332	Sequence 332, App
36	48	21.1	720	3	US-09-257-799-48	Sequence 48, Appl
37	47.5	20.9	98	2	US-08-920-919A-48	Sequence 3, Appl
38	47.5	20.9	98	2	US-08-918-727-3	Sequence 3, Appl
39	47.5	20.9	98	3	US-09-205-680A-3	Sequence 3, Appl
40	47.5	20.9	98	3	US-08-948-889-11	Sequence 11, Appl
41	47.5	20.9	270	2	US-08-852-743-5	Sequence 5, Appl
42	47.5	20.9	270	3	US-09-185-370-5	Sequence 2, Appl
43	47.5	20.9	15281	2	US-08-471-119A-2	Sequence 53, Appl
44	47	20.7	92	3	US-08-813-884-53	Sequence 52, Appl
45	47	20.7	93	3	US-08-813-884-52	Sequence 52, Appl

ALIGNMENTS

RESULT 1
US-08-663-566A-13
Sequence 13, Application US/08663566A
Patent No. 5853733
GENERAL INFORMATION:
APPLICANT: Cochran, Mark D
APPLICANT: Macdonald, Richard D
TITLE OF INVENTION: Recombinant Herpesvirus of Turkeys
TITLE OF INVENTION: and Use thereof
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/663,566A
FILING DATE: June 13, 1996
CLASSIFICATION: A35
ATTORNEY/AGENT INFORMATION:
NAME: White, John P
REGISTRATION NUMBER: 28,678
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)278-0400
TELEFAX: (212)391-0526
TELEX: 422523
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 553 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-663-566A-13

Query Match 24.4%; Score 55.5; DB 2; Length 553;
Best Local Similarity 33.3%; Pred. No. 3.4;
Matches 16; Conservative 7; Mismatches 20; Indels 5; Gaps 1;

4 TSLKNDKFEKITDELEITIGGSLSTFFLENRSTQALGK 46
DB 433 TRLSGEDVYVKNIISDQVITGTMDISTELGNVNSISALAK 480


```

RESULT      5          US-08-362-240A-13
; Sequence 13; Application US/08362240A
; Patent No. 3965138
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D
; APPLICANT: Junker, David
; APPLICANT: Wild, Martha A
; TITLE OF INVENTION: Recombinant Herpesvirus and Uses Thereof
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,240A
; FILING DATE: Dec-22-'94
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)278-0400
; TELEFAX: (212)391-0526
; TELEX: 422523
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-362-240A-13

Query Match      24.4%; Score 55.5; DB 2; Length 553;
Best Local Similarity 33.3%; Pred. No. 3.4;
Matches 16; Conservative 7; Mismatches 20; Indels 5; Gaps 1.

Oy      4 TSLKNDP-----KEIKTDELEITIGSGSSLSFFRLFRSTOLGK 46
        |||:|||:|||||:|||||:|||||:|||||:|||||:|||||
Db      433 TURLSGEPDYVYQKNISTQDSOVITIGIDISTELGNVNNSISLNK 480

RESULT      6          US-08-477-459-14
; Sequence 14; Application US/08477459
; Patent No. 6001369
; GENERAL INFORMATION:
; APPLICANT: Mark D. Cochran
; TITLE OF INVENTION: Recombinant Fowlpox Viruses and Uses
; TITLE OF INVENTION: Theoreif
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

	SOURCE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:	
APPLICATION NUMBER:	US/08/477,459
FILING DATE:	07-JUN-1995
CLASSIFICATION:	424
ATTORNEY/AGENT INFORMATION:	
NAMES:	White Esq, John P
TELECOMMUNICATION INFORMATION:	
TELEPHONE:	(212) 278-0400
TELEFAX:	(212) 391-0525
INFORMATION FOR SEQ ID NO:	14:
SEQUENCE CHARACTERISTICS:	
LENGTH:	553 amino acids
TYPES:	amino acid
TOPOLOGY:	linear
MOLECULE TYPE:	protein
US-08-477-459-14	
Query Match	24.4%; Score 55.5; DB 3; Length 553;
Best Local Similarity	33.3%; Pred. No.3.4;
Matches	16; Conservative 7; Mismatches 20; Indels 5; Gaps 1.
OY 4 TSLKNDP----KEIKTDELTETIGSGSSTFFRLNRSFTQALG 46	: : : : :
Dd 433 TLRLSEPDVTVOKNISIODSQVIITGNIDISTELGVNNNSISNALNK 480	
RESULT 7	
: Sequence 14, Application US/08479869	
: Patent No. 6123949	
: GENERAL INFORMATION:	
: APPLICANT: Cochran Ph.D, Mark D	
: TITLE OF INVENTION: Recombinant Fowlpox Virus S-FPV-043 and	
: TITLE OF INVENTION: Uses Thereof	
: NUMBER OF SEQUENCES: 20	
: CORRESPONDENCE ADDRESSES:	
: ADDRESSSEE: John P. White	
: STREET: 30 Rockefeller Plaza	
: CITY: New York	
: STATE: New York	
: COUNTRY: USA	
: ZIP: 10112	
: COMPUTER READABLE FORM:	
: MEDIUM TYPE: floppy disk	
: COMPUTER: IBM PC compatible	
: OPERATING SYSTEM: pc-dos/ms-dos	
: SOFTWARE: Patentin Release #1.0, Version #1.25	
: CURRENT APPLICATION DATA:	
: APPLICATION NUMBER: US/08/479,869	
: FILING DATE:	
: CLASSIFICATION: 424	
: PRIOR APPLICATION DATA:	
: APPLICATION NUMBER: 08/024,156	
: FILING DATE: 26-FEB-1993	
: ATTORNEY/AGENT INFORMATION:	
: NAMES: White Esq, John P	
: TELECOMMUNICATION INFORMATION:	
: TELEPHONE: (212)977-9550	
: TELEFAX: (212)664-0525	
: TELEX: 422523	
: INFORMATION FOR SEQ ID NO: 14:	
: SEQUENCE CHARACTERISTICS:	
: LENGTH: 553 amino acids	
: TYPES: amino acid	
: TOPOLOGY: linear	
: MOLECULE TYPE: protein	
: US-08-479-869-14	
Query Match	24.4%; Score 55.5; DB 3; Length 553;
Best Local Similarity	33.3%; Pred. No.3.4;

STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/02252A
FILING DATE: 28-FEB-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: White Esq, John P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)977-9550
TELEFAX: (212)664-0525
TELEX: 422523
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 553 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-02252A-14

Query Match 24.4%; Score 55.5; DB 5; Length 553;
Best Local Similarity 33.3%; Pred. No. 3.4;
Matches 16; Conservative 7; Mismatches 20; Indels 5; Gaps 1;

OY 4 TLISKDNF-----KEIKTDELEITIGGSGSLSTFFRLFNRSFTQALGK 46
DB 433 TLRISGEFDVYQKNISIDQSQVITITGNLDISTELGNVNSISNALNK 480

RESULT 12

PCT-US95-10245-13

Sequence 13, Application PC/TUS9510245
GENERAL INFORMATION:
APPLICANT: SYNTRO CORPORATION

TITLE OF INVENTION: Recombinant Herpesvirus of Turkeys And Uses Thereof
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White

STREET: 1185 Avenue of the Americas
CITY: New York

STATE: New York
COUNTRY: USA

ZIP: 10036

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10245

FILING DATE: 09-AUG-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:

NAME: White, John P
REGISTRATION NUMBER: 28,678

TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)278-0400

TELEFAX: (212)391-0526
TELEX: 422523

INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:

LENGTH: 553 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein

PCT-US95-10245-13

Query Match 24.4%; Score 55.5; DB 5; Length 553;
Best Local Similarity 33.3%; Pred. No. 3.4;
Matches 16; Conservative 7; Mismatches 20; Indels 5; Gaps 1;

OY 4 TLISKDNF-----KEIKTDELEITIGGSGSLSTFFRLFNRSFTQALGK 46
DB 433 TLRISGEFDVYQKNISIDQSQVITITGNLDISTELGNVNSISNALNK 480

RESULT 13

5310678-1

Patent No. 5310678
APPLICANT: Bingham, Richard W.; Chambers, Phillip; Emerson, Peter

T. Miller, Neil S.

TITLE OF INVENTION: NEWCASTLE DISEASE VIRUS GENE CLONES
NUMBER OF SEQUENCES: 3

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/438,945

FILING DATE: 17-NOV-1989
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 885,765
FILING DATE: 15-JUL-1986

SEQ ID NO: 1:
LENGTH: 553

5310678-1

Query Match 24.4%; Score 55.5; DB 6; Length 553;
Best Local Similarity 33.3%; Pred. No. 3.4;
Matches 16; Conservative 7; Mismatches 20; Indels 5; Gaps 1;

OY 4 TLISKDNF-----KEIKTDELEITIGGSGSLSTFFRLFNRSFTQALGK 46
DB 433 TLRISGEFDVYQKNISIDQSQVITITGNLDISTELGNVNSISNALNK 480

RESULT 14

US-09-564-805-228

Sequence 228, Application US/09564805
Patent No. 5333403
GENERAL INFORMATION:

APPLICANT: Tavligian, Sean V.
APPLICANT: Teng, David H.F.

APPLICANT: Simard, Jacques
APPLICANT: Rommens, Johanna M.

APPLICANT: Myriad Genetics, Inc.
TITLE OF INVENTION: Chromosome 17p-linked Prostate Cancer Susceptibility

FILE REFERENCE: 2318-258
TITLE OF INVENTION: Gene and a Paralog and Orthologous Genes

CURRENT APPLICATION NUMBER: US/09/564,805
CURRENT FILING DATE: 2000-05-05

PRIOR APPLICATION NUMBER: US 60/107,468
PRIOR FILING DATE: 1998-11-06

PRIOR APPLICATION NUMBER: 09/434,382
PRIOR FILING DATE: 1999-11-05

NUMBER OF SEQ ID NOS: 240
SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 228
LENGTH: 837

TYPE: PRT
ORGANISM: Arabidopsis thaliana

US-09-564-805-228

Query Match 24.4%; Score 55.5; DB 4; Length 837;
Best Local Similarity 34.1%; Pred. No. 5.8;
Matches 14; Conservative 9; Mismatches 9; Indels 9; Gaps 1;

OY 7 LKNDKEIKTDELEITIGGSGS-----LSTFFRLFNK 38
DB 473 LPNCLEKIRDDMEIVILIGSSGSPSKYRNVAIFIDFSR 513

RESULT 15
US-09-242-690A-6
; Sequence 6, Application US/09242690A
; Patent No. 6284534
; GENERAL INFORMATION:
; APPLICANT: KONDO, KEIJI
; APPLICANT: MIURA, YUTAKA
; TITLE OF INVENTION: YEAST VECTOR AND METHOD OF PRODUCING PROTEINS USING THE
; TITLE OF INVENTION: SAME
; FILE REFERENCE: 049441/0118
; CURRENT APPLICATION NUMBER: US/09/242,690A
; CURRENT FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: PCT/JP97/02924
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: JP 8/241062
; PRIOR FILING DATE: 1996-08-23
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 97
; TYPE: PRT
; ORGANISM: Candida utilis
US-09-242-690A-6

Query Match 23.8%; Score 54; DB 4; Length 97;
Best Local Similarity 41.48; Pred. NO. 0.59;
Matches 12; Conservative 6; Mismatches 11; Indels 0; Gaps 0;
QY 1 MKKTLKNDPEKIKTDELEIIIGSGSL 29
|||: ||: ||| | : : |
Db 43 MKKTIYENGFRKIKGYQLVYVNSDKL 71

Search completed: July 30, 2002, 15:14:41
Job time: 517 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:08:49 ; Search time 29.21 Seconds

(without alignments)
151.322 Million cell updates/sec

Title: US-09-833-017-2

Perfect score: 227
Sequence: 1 MKKTLKNDPKEIKTDELE.....GSLSTFFRLNRSFTQALGK 46

Scoring table: BLOSUM62
Gapop 10.0 , Gapect 0.5

Searched: 283138 segs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	69	30.4	850	2	ABC transporter, p
2	69	30.4	850	2	conserved hypochet
3	60.5	26.7	475	2	hypothetical prote
4	58.5	25.8	41	2	competence stimula
5	58	25.6	1009	2	hypothetical prote
6	57.5	23.3	516	2	glucose-6-phosphat
7	56.5	22.9	398	2	betaine betaine/1-
8	56.5	24.9	408	2	ABC transporter AR
9	56.5	24.9	593	2	ABC transporter AR
10	56.5	24.9	893	2	pyrroline-5-carbox
11	56	24.7	893	2	pyrroline-5-carbox
12	56	24.7	1250	2	probable restricti
13	55.5	24.4	416	2	probable NAD(P)-de
14	55.5	24.4	416	2	probable NAD(P)-de
15	55.5	24.4	553	1	cell fusion glycop
16	55.5	24.4	553	1	cell fusion glycop
17	55.5	24.4	553	1	cell fusion glycop
18	55.5	24.4	553	1	cell fusion glycop
19	55.5	24.4	553	1	cell fusion glycop
20	55.5	24.4	553	1	cell fusion glycop
21	55	24.2	278	2	unknown protein [l
22	55	24.2	278	2	hypothetical prote
23	55	24.2	278	2	hypothetical prote
24	54.5	24.0	608	2	centromere protein
25	54.5	24.0	330	2	cell-division ATP-
26	54.5	24.0	361	2	opuna (AF234619) [
27	54.5	24.0	361	2	histidinol-phospha
28	54.5	24.0	463	2	chromosomal replic
29	54.5	24.0	720	2	phenylalanine--trn

30	54	23.8	226	1	C49769	gap junction prote
31	54	23.8	226	1	A33646	gap junction prote
32	54	23.8	243	2	C96946	glutamine ABC tran
33	54	23.8	708	2	T24727	hypothetical prote
34	54	23.8	1292	2	F64237	DNA-directed RNA p
35	53.5	23.6	165	2	S16722	coenzyme F420 hyd
36	53.5	23.6	241	2	AC1356	amino acid ABC tra
37	53.5	23.6	241	2	AD1726	amino acid ABC tra
38	53.5	23.6	242	2	AH1180	amino acid ABC tra
39	53.5	23.6	242	2	A11537	amino acid ABC tra
40	53.5	23.6	243	2	D97008	ABC-type polar ami
41	53.5	23.6	246	2	G81409	probable phosphate
42	53.5	23.6	331	2	H81300	probable anion-upt
43	53.5	23.6	553	1	VGN201	cell fusion glycop
44	53.5	23.6	553	1	A46329	cell fusion glycop
45	53.5	23.6	553	1	B46329	cell fusion glycop

ALIGNMENTS

RESULT 1
F95260
ABC transporter, permease protein, probable SP2231 [Imported] - Streptococcus pneumon
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: F95260
R:Retelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzappl
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: F95260
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-850 <KUR>
A:Cross-references: GB:AE005672; FIDN:AAK76279.1; PID:g14973742; GSPDB:GN00164; TIGR:
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP2231

Query Match 30.4%; Score 69; DB 2; Length 850;
Best Local Similarity 33.3%; Pred. No. 1.1;
Matches 19; Conservative 10; Mismatches 16; Indels 12; Gaps 2;

QY 2 KKTLSLKNDP-----KEIKTDELEITIGSGSLST-----TFRRLNRSFTQALGK 46

DB 692 KKTFTTDVNFSPFNIGYTKKTKENINVSFPGNSQVSFSEPTFRLDTKTFETAIQK 748

RESULT 2
A98126
Conserved hypothetical protein spr2036 [Imported] - Streptococcus pneumoniae (strain

C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001

C:Accession: A98126

R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.
y, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.

J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.

A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain K6.

A:Reference number: A97872; MUID:21429245; PMID:11544234

A:Accession: A98126

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-850 <KUR>

A:Cross-references: GB:AE007317; FIDN:ALU00838.1; PID:g15459743; GSPDB:GN00174

C:Genetics:
A:Gene: spr2036

```

Query Match      30.4%; Score 69; DB 2; Length 850;
Best Local Similarity 33.3%; Pred. No. 1,1;
Matches 19; Conservative 10; Mismatches 16; Indels 12; Gaps 2;

QY          2   KKTSLKNDP-----KEIKTDELEIIIGSGSLS---TFPRLFNRSTQALGK 46
              ||| : | |           ||| : : : : | | :||| :||| :||| :|
Db    692   KKTFTTDVNFSEFMIGTYKEKKTFENINVSFPGNSQVSEESPTFYRLDTKTTETAIOK 748

RESULT       3
S52893        hypothetical protein YMR044w - yeast (Saccharomyces cerevisiae)
N5Altername alternate names: hypothetical protein YM9532..09
C5Species: Saccharomyces cerevisiae
C5Date: 08-May-1995 #sequence_revision 01-Sep-1995 #text_change 05-Nov-1999
C5Accession: S52893
R5OdeLL,C.; Bowman, S.
submitted to the EMBL Data Library, February 1995
A5Reference number: S52885
A5Accession: S52893
A5Molecule type: DNA
A5Residues: 1-475 <CODE>
A5Cross-references: EMBL,Z48502; NID:9695715; PIDN:CAA88410.1; PID:9695724; GSPDB:GN00001
C5Genetics:
A5Gene: MIPS:YMR044w
A5Map position: 13R
```

Query Match	26.7%	Score 60.5	DB 2	length 475
Best Local Similarity	38.9%	Pred. No. 6.9		
Matches 14	Conservative 9	Mismatches 12	Indels 1	Gaps 1
QY	1 MKRTSLKNDKFKETDELEITIGGSGSLSTFRRLF	36		
	: : : : : :			
Db	361 INQAVSIRKDF-ETILDELQIALDTRGSRNEETITTF	395		

RESULT 4
H98126 competence stimulating peptide precursor (CSP) [imported] - Streptococcus pneumoniae (st
C.Species: Streptococcus pneumoniae
C.Date: 22-oct-2001 #sequence_revision 22-oct-2001 #text_change 22-oct-2001
C.Accession: H98126
R.Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
y, R.; LeBlanc, D.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A.Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A.Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A.Reference number: A97872; MUID:21429245; PMID:11544234
A.Accession: H98126
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-41 <KUR>
A.Cross-references: GB:AE007317; PIDN:AAL00845.1; PID:g15459751; GSPDB:GN00174
C.genetics:
A:gene: comC

Query Match	25.8%	Score 58.5	DB 2	Length 41
Best Local Similarity	44.4%	Pred. No. 0.83		
Matches 16	Conservative 5	Mismatches 14	Indels 1	Gaps 1

RESULT 5
E89792
hypothetical protein SA0272 [imported] - Staphylococcus aureus (strain N315)

C:Species: *Staphylococcus aureus*
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: E89792
R:Kunoda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, ma, A.; Mizutani-Oli, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Seki, C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratsugu, K.
L:ncet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
A:Reference number: AB9758; MUID:21311952; PMID:11418146
A:Accession: E89792
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1009 <RUR>
A:Cross-references: GB:BA000018; PID:g13700198; PIDN:BAB41496.1; GSEDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: SA0272

Query Match	25.6%	Score 58	DB 2	Length 1009
Best Local Similarity	34.3%	Pred. NO. 34		
Matches	12	Conservative	8	Mismatches 15; Indels 0; Gaps 0;
QY	9	NDFKEIKTDELEIIIGSGSLSTFFRLFNRSFTQA	43	
DB	202	NDFPELFTDLVNSISANKDITKMFOTYKSLISA	236	

```

RESULT      6
glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49) [imported] - Arabidopsis thaliana
T52611
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 03-Nov-2000
C:Accession: T52611
R:Rent, U.K.; Hauschild, R.; Lange, C.; Pieferma, M.; Wenderoth, I.; von Schaeven,
Plant Mol. Biol. 40, 487-494, 1999
A>Title: Evidence for functional convergence of redox regulation in G6PDH isoforms of
A:Reference number: 225284
A:Accession: T52611
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-516 <KRN>
A:Cross references: EMBL:AJ010970; PDB: CAB52674.1
C:Genetics:
C:Gene: acg9
C:Superfamily: glucose-6-phosphate dehydrogenase
C:Keywords: oxidoreductase

```

```

Query Match 25.3%; Score 57.5; DB 2; Length 516;
Best Local Similarity 41.5%; Pred. No. 19;
Matches 22; Conservative 7; Mismatches 11; Indels 13; Gaps 5;

OY 1 MKTLLSLKND--EKE-----IKTDELEIIT--GGSGSIS---TF---FRINRSP 40
   1 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 8 MEKSTSLTKNDSEVKEEYNPITETGSLSLTLAGASDGLAKKTFPALNINRQCF 60
   1 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 7
 AC3578
 glycine betaine/L-proline transport ATP-binding protein prov BMEI10548 [imported] - B
 C:Species: *Brucella melitensis*
 C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
 C:Accession: AC3578
 R:DelVecchio, V.G.; Kapatakar, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanov
 I.; Mazur, M.; Goldsman, E.; Selakos, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Lett
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
 A:title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*
 A:Reference number: AD3252; PMID:11756688

A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-398 <KUR>

A:Cross-references: GB:AE008918; PIDN:AAL53790.1; PID:g17984720; GSPDB:GN00191
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI0548
A:Map position: II

Query Match 24.9%; Score 56.5; DB 2; Length 398;
Best Local Similarity 37.2%; Pred. No. 19;
Matches 16; Conservative 7; Mismatches 19; Indels 1; Gaps 1;

OY 4 TLCLKNDKEIKTDELEIIGSGS-LSTFFRLFNRSFTQALG 45
DB 55 TIGVDNATFDIREGEVFMGLSSGSKSTLRLLNRLIEPTAG 97

RESULT 8
D86806
betaine ABC transporter ATP binding protein busAA [imported] - Lactococcus lactis subsp.
C:Species: Lactococcus lactis subsp. lactis
C>Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: D86806
R:Botolin, A.; Wincker, P.; Mauge, S.; Jaillon, O.; Malarme, K.; Weissbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss
A:Reference number: A86625; MUID:11235186; PMID:11337471
A:Accession: D86806
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-408 <STO>
A:Cross-references: GB:AE005176; PID:g12724443; PIDN:AAK05550.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: busAA

Query Match 24.9%; Score 56.5; DB 2; Length 408;
Best Local Similarity 45.7%; Pred. No. 19;
Matches 16; Conservative 3; Mismatches 15; Indels 1; Gaps 1;

OY 13 EIKTDELEIIGSGS-LSTFFRLFNRSFTQALG 46
DB 50 EINEGEIVIMGLSSGSKSTLRLLNRLIEPTSGK 84

RESULT 9
C97848
ABC transporter ATP-binding protein uup [imported] - Rickettsia conorii (strain Malish 7
C:Species: Rickettsia conorii
C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
C:Accession: C97848
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro
Science 293, 2093-2098, 2001
A:Title: Mechanisms of evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MUID:21442074; PMID:11557893
A:Accession: C97848
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-593 <KUR>
A:Cross-references: GB:AE006914; PIDN:AAL03725.1; PID:g15620316; GSPDB:GN00173
C:Genetics:
A:Gene: uup
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 24.9%; Score 56.5; DB 2; Length 593;
Best Local Similarity 37.0%; Pred. No. 29;
Matches 17; Conservative 5; Mismatches 23; Indels 1; Gaps 1;

OY 2 KKTSLKNDKEIKTDELEIIGSGS-LSTFFRLFNRSFTQALG 46
DB 298 KMTKINNPFVKNGEKIIGANGSGSKTFIKLLTKQLTPESGK 343

RESULT 10
S64926
probable membrane protein YLR092W - Yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein L2528
C:Species: Saccharomyces cerevisiae
C>Date: 01-Aug-1995 #sequence_revision 24-May-1996 #text_change 06-Feb-1998
C:Accession: S64926
R:Benes, V.; Reichmann, S.; Nentwich, U.; Schwager, C.; Ansoorge, W.; Voss, H.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64920
A:Accession: S64926
A:Molecule type: DNA
A:Residues: 1-893 <BEN>
A:Cross-references: EMBL:Z73264; NID:g1360466; PID:e245799; PID:g1360467; MIPS:YLR092

A:Experimental source: strain S288C
C:Genetics:
A:Gene: SCD:SUL2
A:Cross-references: SCD:S0004082; MIPS:YLR092W
A:Map position: 12R
C:Keywords: transmembrane protein
F:136-152/Domain: transmembrane #status predicted <TM1>
F:223-239/Domain: transmembrane #status predicted <TM2>
F:244-260/Domain: transmembrane #status predicted <TM3>
F:302-318/Domain: transmembrane #status predicted <TM4>
F:358-374/Domain: transmembrane #status predicted <TM5>
F:411-427/Domain: transmembrane #status predicted <TM6>
F:488-504/Domain: transmembrane #status predicted <TM7>
F:538-554/Domain: transmembrane #status predicted <TM8>
F:557-573/Domain: transmembrane #status predicted <TM9>

Query Match 24.9%; Score 56.5; DB 2; Length 893;
Best Local Similarity 36.7%; Pred. No. 46;
Matches 18; Conservative 6; Mismatches 18; Indels 7; Gaps 2;

OY 1 MKKTSLKNDKEIKTDELEIIGSGSLSTFFRL-----FKRSTQA 43
DB 426 ISKSPGRINDYKVPVDEL-IAIGVSNLGTFFNAYPATGFSRSALKA 473

RESULT 11
T12258
pyrroline-5-carboxylate synthetase (PC 1.5.1.1) - common ice plant
C:Species: Mesembryanthemum crystallinum (common ice plant)
C>Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 17-Nov-2000
C:Accession: T12258
R:Michalowski, C.B.; Ouilgley-Landreau, F.; Bohner, H.J.
submitted to the EMBL data library, May 1998
A:Description: Mesembryanthemum crystallinum pyrroline-5-carboxylate synthetase mRNA.
A:Reference number: Z17475
A:Accession: T12258
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-719 <MTC>
A:Cross-references: EMBL:AF067967; NID:g3176964; PID:g3176965
C:Genetics:
A:Gene: P5CS
C:Superfamily: glutamate-5-semialdehyde dehydrogenase
C:Keywords: oxidoreductase

Query Match 24.7%; Score 56; DB 2; Length 719;
Best Local Similarity 36.6%; Pred. No. 42;
Matches 15; Conservative 7; Mismatches 13; Indels 6; Gaps 1;

OY 6 SLKNDKEIKTDELEIIGSGSLST-----PFRFNSE 40
DB 37 SLCEOLKELNSDGYEVILVTSAGVAGRORLRFKRLVNSSF 77

RESULT 12
E81339

probable restriction/modification enzyme Cj0690c [imported] - *Campylobacter jejuni* (str C:Species: *Campylobacter jejuni*
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Mar-2000
 C:Accession: E81339
 R:Parhill, J.; Wren, B.W.; Mungall, K.; Kelley, J.M.; Churcher, C.; Basham, D.; Chillingworth, C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVlier, A.; Whitehead, S.; Barré-Nature 403, 665-668, 2000
 A:Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hlyA
 A:Reference number: AB1250; MUID:20150912
 A:Accession: E81339
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1250 <PAR>
 A:Cross-references: GB:AL139076; GB:AL111168; NID:g6968128; PIDN:CAB72964.1; PID:g696814
 A:Experimental source: serotype O2, strain NCTC 11168
 C:Genetics:
 A:Gene: Cj0690c

Query Match 24.4%; Score 55.5; DB 2; Length 1250;
 Best Local Similarity 40.0%; Pred. No. 77;
 Matches 14; Conservative 4; Mismatches 15; Indels 2; Gaps 1;

OY 3 KTLSTKNDKE-1KTDELEIIIGSGSLSTFRLFN 37
 DB 380 KNLSEKRDYKTLSEVHLGTY--EGLISTFEEIAN 412

RESULT 13
 AC1143
 probable NAD(P)-dependent oxidoreductase homolog lmo0546 [imported] - *Listeria monocytog*

C:Species: *Listeria monocytogenes*
 C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
 C:Accession: AC1143
 R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
 .; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurgey, O.; Entian, K.D.; Fsihl, H.
 D.; Jones, L.M.; Karst, U.
 Science 294, 849-852, 2001
 A:Authors: Krefl, U.; Kuhn, M.; Kunst, F.; Kurapk, G.; Madueno, E.; Maitournam, A.; Ma
 ok, C.; Schluter, T.; Simoes, N.; Tietz, A.; Vazquez-Boland, J.A.; Voss, H.; Weiland,
 A:Title: Comparative genomics of *Listeria species*.
 A:Reference number: AB1077; MUID:21537279; PMID:11679669
 A:Accession: AC1143
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-416 <GLA>
 A:Cross-references: GB:NC_003210; PIDN:CAC98625.1; PID:g16409922; GSPDB:GN00177
 A:Experimental source: strain EGD-e
 C:Genetics:
 A:Gene: lmo0546

Query Match 24.4%; Score 55.5; DB 2; Length 416;
 Best Local Similarity 44.4%; Pred. No. 26;
 Matches 12; Conservative 7; Mismatches 7; Indels 1; Gaps 1;

OY 2 KKTSLKNDKE-1KTDELEIIIGSG 27
 DB 68 KETILSNDFKEIHSDELVEIVDATG 94

RESULT 14
 AF1501
 probable NAD(P)-dependent oxidoreductase homolog lln0550 [imported] - *Listeria innocua*

C:Species: *Listeria innocua*
 C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
 C:Accession: AF1501
 R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
 .; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurgey, O.; Entian, K.D.; Fsihl, H.
 D.; Jones, L.M.; Karst, U.
 Science 294, 849-852, 2001
 A:Authors: Krefl, U.; Kuhn, M.; Kunst, F.; Kurapk, G.; Madueno, E.; Maitournam, A.; Ma
 ok, C.; Schluter, T.; Simoes, N.; Tietz, A.; Vazquez-Boland, J.A.; Voss, H.; Weiland,

A:Title: Comparative genomics of *Listeria species*.
 A:Reference number: AB1077; MUID:21537279; PMID:11679669
 A:Accession: AF1501
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-416 <GLA>
 A:Cross-references: GB:AL592022; PIDN:CAC95782.1; PID:g16412990; GSPDB:GN00178
 A:Experimental source: strain C1p11262
 C:Genetics:
 A:Gene: lln0550

Query Match 24.4%; Score 55.5; DB 2; Length 416;
 Best Local Similarity 44.4%; Pred. No. 26;
 Matches 12; Conservative 7; Mismatches 7; Indels 1; Gaps 1;

OY 2 KKTSLKNDKE-1KTDELEIIIGSG 27
 DB 68 KETILSNDFKEIHSDELVEIVDATG 94

RESULT 15
 VGZNV
 cell fusion glycoprotein precursor - Newcastle disease virus (strains Beaudette C and

N:Contains: fusion glycoprotein F1; fusion glycoprotein F2
 C:Species: Newcastle disease virus
 C:Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 16-Jul-1999
 C:Accession: A27008; F46329
 R:Chambers, P.; Miller, N.S.; Emerson, P.T.
 J. Gen. Virol. 67, 2685-2694, 1986
 A:Title: Nucleotide sequence of the gene encoding the fusion glycoprotein of Newcastle
 A:Reference number: A27008; MUID:87085486
 A:Accession: A27008
 A:Molecule type: mRNA
 A:Residues: 1-553 <CH>
 A:Cross-references: GB:XA4719; NID:g60933; PIDN:CAA28426.1; PID:g60934
 A:Experimental source: strain Beaudette C
 R:Toyoda, T.; Sakaguchi, T.; Hirota, H.; Gotoh, B.; Kuma, K.; Miyata, T.; Nagai, Y.
 Virology 169, 273-282, 1989
 A:Title: Newcastle disease virus evolution. II. Lack of gene recombination in generat
 A:Reference number: A46329; MUID:89204898

A:Accession: F46329
 A:Molecule type: genomic RNA
 A:Residues: 1-553 <TOY>
 A:Cross-references: GB:M24697; NID:g293929; PIDN:AAA46648.1; PID:g293930
 A:Experimental source: strain BEA/45
 C:Genetics:
 A:Gene: F

C:Superfamily: paraInfluenza virus cell fusion protein
 C:Keywords: glycoprotein; membrane fusion; transmembrane protein
 F:1-25/Domain: signal sequence #status predicted <SIG>
 F:26-117/Product: cell fusion glycoprotein F2 #status predicted <FP2>
 F:118-553/Product: cell fusion glycoprotein F1 #status predicted <FP1>
 F:495-528/Domain: transmembrane #status predicted <TMN>
 F:85,191,366,447/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 24.4%; Score 55.5; DB 1; Length 553;
 Best Local Similarity 33.3%; Pred. No. 36;
 Matches 16; Conservative 7; Mismatches 20; Indels 5; Gaps 1;

OY 4 TISLKNDF-----KEIKTDELEIIIGSGSLSTFRLFRSFTALGK 46
 DB 433 TRLSGEDFATYOKNISIDSOVITIGNDISTELGNVNNISNALNK 480

Search completed: July 30, 2002, 15:15:18
 Job time: 389 sec

Wed Jul 31 08:30:53 2002

us-09-833-017-2.rpt

Page 5

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:14:49 ; Search time 15.86 seconds
(without alignments)
112.301 Million cell updates/sec

Title: US-09-833-017-2
227
Sequence: 1 MKRTSLKNDKFEIKTDELE.....GSLSTFFRLFNRSFTQALGK 46

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	60.5	26.7	475 1 YMTA_YEAST	004213 saccharomyc
2	58.5	25.8	41 1 CSPL_STRPN	054712 streptococ
3	56.5	24.9	893 1 SOL2_YEAST	012325 saccharomyc
4	56	24.7	719 1 P5C5_MESCR	065361 m delta 1-p
5	55.5	24.4	41 1 CSP_STROR	033689 streptococ
6	55.5	24.4	553 1 VGLF_NDVB	P06156 newcastie d
7	55.5	24.4	553 1 VGLF_NDVL	P33613 newcastie d
8	55.5	24.4	553 1 VGLF_NDVL	P33614 newcastie d
9	55.5	24.4	553 1 VGLF_NDVL	P12571 newcastie d
10	55	24.2	608 1 CB32_YEAST	P40969 saccharomyc
11	54.5	24.0	373 1 HTS8_METUA	Q58365 methanococ
12	54.5	24.0	463 1 DNAA_RICCN	Q92856 rickettsia
13	54.5	24.0	720 1 SYFB_PORPU	P51346 porphyra pu
14	54	23.8	226 1 CXB2_MOUSE	Q00977 mus musculu
15	54	23.8	226 1 CXB2_MOUSE	P21994 rattus norv
16	54	23.8	708 1 YNZB_CARFL	P45972 caenorhabdi
17	54	23.8	1292 1 RPOC_MYCGE	P47582 mycoplasma
18	53.5	23.6	259 1 PSTB_EDMTA	Q9am14 mycoplasma
19	53.5	23.6	553 1 VGLF_NDVB	P33936 newcastie d
20	53.5	23.6	553 1 VGLF_NDVB	P33615 newcastie d
21	53.5	23.6	553 1 VGLF_NDVB	P12570 newcastie d
22	53.5	23.6	553 1 VGLF_NDVB	P12570 newcastie d
23	53	23.3	274 1 PNK_AOUAE	O67055 aquifex aeo
24	53	23.3	400 1 PROV_ECOTI	P14175 escherichia
25	53	23.3	400 1 PROV_ECOTI	P17328 salmonella
26	53	23.3	513 1 LEU1_LACLA	Q02141 lactococcus
27	53	23.3	513 1 LEU1_LACLA	Q02141 lactococcus
28	53	23.3	1132 1 DNBI_HSV6U	P42846 saccharomyc
29	53	23.3	1132 1 DNBI_HSV6U	P52538 human herpe
30	52.5	23.1	310 1 GPRB_BP22	P57022 bacillopho
31	52.5	23.1	564 1 V61A_MYCPN	P75041 mycoplasma
32	51.5	22.7	41 1 CSP2_STROR	033689 streptococ
33	51.5	22.7	240 1 PPRH_METUA	Q58656 methanococ

34	51.5	22.7	252 1 Y830_METUA	Q58240 methanococ
35	51.5	22.7	257 1 PSTB_ECOTI	P07655 escherichia
36	51.5	22.7	259 1 DNAA_MYCBO	P49991 mycobacteri
37	51.5	22.7	306 1 GTRB_ECOTI	P77293 escherichia
38	51.5	22.7	306 1 GTRB_ECOTI	P03866 staphylococ
39	51	22.5	295 1 UNG_MOUSE	P97931 mus musculu
40	51	22.5	304 1 UNG_MOUSE	P13051 homo sapien
41	51	22.5	360 1 DERA_DIACA	P51104 dianthus ca
42	51	22.5	397 1 Y636_METUA	Q58053 methanococ
43	51	22.5	461 1 Y061_MYCGE	P47307 mycoplasma
44	51	22.5	465 1 CG12_CANAL	P43062 candida alb
45	51	22.5	511 1 G6PD_EMENTI	P41764 emeticellia

ALIGNMENTS

RESULT 1
YMTA_YEAST STANDARD; PRT; 475 AA.

AC Q04213;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 55.4 kDa protein in MCML-NDP16 intergenic region.
GN YMR044W OR YMR532.09.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C / AB972;
RA Odell C., Bowman S., Barrell B.G., Rajandream M.A.;
RL Submitted (Feb-1995) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation CC between the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb.ch/announce/> or send an email to license@isb-sib.ch).

CC
DR EMBL; 248502; CAA8410.1; -
DR SGD; S0004647; YMR044W.
DR InterPro: IPR000313; PWM.
DR Pfam: PF00855; PWM; 1.
DR SMART; SM00293; PWM; 1.
KW Hypothetical protein.
KW DOMAIN 74
FT DOMAIN 186 195
FT DOMAIN 243 251
FT POLY-GLU.
FT POLY-GLU.
SQ SEQUENCE 475 AA; 55426 MW; DFD6189E81178060 CRC64;

Query Match 26.7%; Score 60.5; DB 1; Length 475;
Best Local Similarity 38.9%; Pred. No. 2.8;
Matches 14; Conservative 9; Mismatches 12; Indels 1; Gaps 1;

OY 1 MKRTSLKNDKFEIKTDELEITIGGSLSTFFRLF 36
DB 361 INQAVSINKDF-ELILDELTALDTRGSRNFEITIF 395

RESULT 2
ID CSPL_STRPN STANDARD; PRT; 41 AA.
AC Q54712;

DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Competence stimulating peptide type 1 precursor (CSP-1).

COMC1 OR COMC.
 OS Streptococcus pneumoniae.
 CC Bacteria: Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 CC Streptococcus.
 CC NCBI_TaxID=1313;
 RX MEDLINE=96074663; PubMed=7479953;
 RA Haavarsstein L.S., Coomaraswamy G., Morrison D.A.;
 RT "An unmodified heptadecapeptide pheromone induces competence for
 genetic transformation in Streptococcus pneumoniae.";
 RT Proc. Natl. Acad. Sci. U.S.A. 92:11140-11144(1995).
 RN [2]
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 25-41.
 RC STRAIN=RX / Cpl200;
 RX MEDLINE=97206147; PubMed=9157240;
 RA Cheng Q., Campbell E.A., Naughton A.M., Johnson S., Masure H.R.;
 RT "The com locus controls genetic transformation in Streptococcus
 pneumoniae.";
 RT Mol. Microbiol. 23:683-692(1997).
 CC -1- FUNCTION: ACTS AS A PHEROMONE, INDUCES CELLS TO DEVELOP COMPETENCE
 FOR GENETIC TRANSFORMATION.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO THE COMC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U33315; AAC44440.1; -;
 DR EMBL: U76218; AAC44895.1; -;
 DR InterPro: IPR004286; Comc.
 DR Pfam: PF03047; Comc; 1.
 KW Pheromone; Competence.
 FT PROPEP 1 24
 FT CHAIN 25 41
 FT SEQUENCE 41 AA; 4971 MW; 1448841498086866 CRC64;
 SO SEQUENCE

Query Match 25.8%; Score 58.5; DB 1; Length 41;
 Best Local Similarity 44.4%; Pred. No. 0.34;
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;
 OY 1 MKKTLISKNDPKREIKDELEITIGSGSLSTFRL 36
 DB 1 MKNTVKLE-QFVALKEDLOKIKGEMRLSKFTRDF 35

RESULT 3
 SUI2_YEAST 3
 ID SUI2_YEAST STANDARD; PRT; 893 AA.
 AC Q12325;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Sulfate permease 2 (High-affinity sulfate transporter 2).
 GN SUI2 OR SEL2 OR YLR092M OR L9449.1.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycaceae;
 CC NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Johnson M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
 RA Favell A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,
 RA Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,
 RA Johnston L., Langston Y., Latreille P., Le T., Mardis E., Menezes S.,

Miller N., Nhan M., Pauley A., Peluso D., Rifkin L., Riles L.,
 RA Tach A., Trevasakis E., Vignati D., Wilcox L., Woldman P., Vandin M.,
 RA Wilson R., Waterston R.;
 RN Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Benes V., Rechmann S., Wentlich U., Schwager C., Ansoerge W., Voss H.;
 RN Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE=97207837; PubMed=9055073;
 RA Cherest H., Davidian J.C., Thomas D., Benes V., Ansoerge W.,
 RA Sureda-Kerjan Y.;
 RT "Molecular characterization of two high affinity sulfate transporters
 in Saccharomyces cerevisiae.";
 RT Genetics 145:627-635(1997).
 CC -1- FUNCTION: HIGH AFFINITY UPTAKE OF SULFATE INTO THE CELL.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE SULFATE PERMEASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U53880; AAB67596.1; -;
 DR EMBL: Z73264; CAA97653.1; -;
 DR EMBL: Z73265; CAA97655.1; -;
 DR EMBL: U53876; AAB67550.1; -;
 DR SGD: S0004082; SUL2.
 DR InterPro: IPR002645; STAS.
 DR InterPro: IPR001902; Sulfate_transp.
 DR Pfam: PF01740; STAS; 1.
 DR Pfam: PF00916; Sulfate_transp; 1.
 DR PROSITE: PS01130; SULFATE_TRANS; 1.
 KW Transport; Transmembrane; Glycoprotein.
 FT TRANSMEM 132 152
 FT TRANSMEM 164 184
 FT TRANSMEM 189 209
 FT TRANSMEM 222 242
 FT TRANSMEM 245 265
 FT TRANSMEM 306 326
 FT TRANSMEM 351 371
 FT TRANSMEM 400 420
 FT TRANSMEM 444 464
 FT TRANSMEM 484 504
 FT TRANSMEM 539 559
 FT TRANSMEM 560 580
 FT SEQUENCE 893 AA; 99650 MW; 67826595AC7C0BF5 CRC64;
 SO SEQUENCE

Query Match 24.9%; Score 56.5; DB 1; Length 893;
 Best Local Similarity 36.7%; Pred. No. 19;
 Matches 18; Conservative 6; Mismatches 18; Indels 7; Gaps 2;
 OY 1 MKKTLISKNDPKREIKDELEITIGSGSLSTFRL-----FNRSFTQA 43
 DB 426 ISMSFGRINDYKVPDDEL-IALGVSNLGTFFNAYPATGSPRSALKA 473

RESULT 4
 P5CS_MESCR 4
 ID P5CS_MESCR STANDARD; PRT; 719 AA.
 AC O65361;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Delta 1-pyridoxine-5-carboxylate synthetase (P5CS) [Includes: Glutamate
 5-kinase (EC 2.7.2.11) (Gamma-glutamyl kinase) (GK); Gamma-glutamyl
 DE phosphate reductase (GPR) (EC 1.2.1.41) (Glutamate-5-semialdehyde

DE dehydrogenase) (glutamyl-gamma-semialdehyde dehydrogenase)].
 GN P5CS.
 OS Mesembryanthemum crystallinum (Common ice plant).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Caryophyllales; Caryophyllales; Alzaceae; Mesembryanthemum.
 NCBI_TaxID=3544;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Michalowski C.B., Ouygley-Landreau F., Bohnert H.J.;
 RT "Mesembryanthemum crystallinum pyruvate-5-carboxylate synthetase
 mRNA.";
 RL Submitted (May-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: P5CS PLAYS A KEY ROLE IN PROLINE BIOSYNTHESIS, LEADING
 CC TO OSMOREGULATION IN PLANTS.
 CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate = ADP + L-glutamate 5-
 CC phosphate.
 CC -1- CATALYTIC ACTIVITY: L-glutamate 5-semialdehyde + phosphate +
 CC NADP(+) = L-gamma-glutamyl 5-phosphate + NADPH.
 CC -1- ENZYME REGULATION: FEEDBACK REGULATED BY PROLINE.
 CC -1- PATHWAY: FIRST AND SECOND STEPS IN PROLINE BIOSYNTHESIS PATHWAY.
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN LEAVES AND IS
 CC INDUCIBLE IN ROOTS SUBJECTED TO SALT STRESS.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE GLUTAMATE 5-
 CC KINASE FAMILY.
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE GAMMA-
 CC GLUTAMYL PHOSPHATE REDUCTASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF067967; AAC18862.1;
 DR Mendei: 29718; Meser:1229;29718.
 DR InterPro: IPR001048; AakInase.
 DR InterPro: IPR002086; Aldehyde_dehydr.
 DR InterPro: IPR000965; GPR.
 DR InterPro: IPR001057; Glut_5_kinase.
 DR Pfam: PF00696; aakInase; 1.
 DR Pfam: PF00171; aldehyd; 1.
 DR PRINTS: PR00474; GLUTAMATE_5_KINASE; 1.
 DR PROSITE: PS00902; GLUTAMATE_5_KINASE; 1.
 DR PROSITE: PS01223; PROA; 1.
 KM Proline biosynthesis; Multifunctional enzyme; Oxidoreductase; NADP;
 KW Transferase; Kinase.
 FT DOMAIN 1 293 GLUTAMATE 5-KINASE.
 FT 294 719 GAMMA-GLUTAMYL PHOSPHATE REDUCTASE.
 FT SEQUENCE 719 AA; 77855 MW; 1510D1AF559961 CRC64;
 SQ
 Query Match 24.7%; Score 56; DB 1; Length 719;
 Best Local Similarity 36.6%; Pred. No. 17;
 Matches 15; Conservative 7; Mismatches 13; Indels 6; Gaps 1;
 QY 6 SLKNDKFKIKTDELEIIIGSGSLST-----FFRLKRSF 40
 DB 37 SLCEQLKEINSDGEYIVLTGSAVSAGRQRLFRKLVNSSF 77
 RESULT 5
 CSP_STROR STANDARD: PRT; 41 AA.
 AC 033689;
 DT 30-MAY-2000 (rel. 39, Created)
 DT 30-MAY-2000 (rel. 39, Last sequence update)
 DT 30-MAY-2000 (rel. 39, Last annotation update)
 DE Competence stimulating peptide precursor (CSP).
 GN COMC.
 OS Streptococcus oralis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus
 OC NCBI_TaxID=1303;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCIC 11427;
 RX MEDLINE=98012953; PubMed=9352904;
 RA Haeverstein L.S., Hakenbeck R., Gaustad P.;
 RT "Natural competence in the genus Streptococcus: evidence that
 RT streptococci can change phenotype by interspecies recombinational
 RT exchanges.";
 RL J. Bacteriol. 179:6589-6594(1997).
 CC -1- FUNCTION: ACTS AS A PHEROMONE, INDUCES CELLS TO DEVELOP COMPETENCE
 CC FOR GENETIC TRANSFORMATION.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO THE COMC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AJ000873; CA04363.1; -
 DR InterPro: IPR004288; COMC.
 DR Pfam: PF03047; COMC; 1.
 KW Pheromone; Competence.
 FT PROPEP 1 24 POTENTIAL.
 FT CHAIN 25 41 COMPETENCE STIMULATING PEPTIDE.
 FT SEQUENCE 41 AA; 4988 MW; 127E5B6F72A1463 CRC64;
 SQ
 Query Match 24.4%; Score 55.5; DB 1; Length 41;
 Best Local Similarity 46.3%; Pred. No. 0.82;
 Matches 19; Conservative 5; Mismatches 14; Indels 3; Gaps 3;
 QY 1 MKKTLISKNDKFKIKTDELEIIIGSGSLSTFFR-LF-NRS 39
 DB 1 MKNTKLE-QFKEVTEAELEIGRSGDKRLPYFFKILFSNRT 40
 RESULT 6
 ID VGLF_NDVV STANDARD: PRT; 553 AA.
 AC P06156;
 DT 01-JAN-1988 (rel. 06, Created)
 DT 01-JAN-1988 (rel. 06, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Fusion glycoprotein precursor [contains: Fusion glycoprotein F2;
 DE Fusion glycoprotein F1].
 GN F.
 OS Newcastle disease virus (strain Beaudette C/45) (NDV).
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
 OC Paramyxoviridae; Paramyxovirinae; Rubulavirinae;
 OC NCBI_TaxID=11178;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87083486; PubMed=3025345;
 RA Chambers P., Millar N.S., Emmerston P.T.;
 RT "Nucleotide sequence of the gene encoding the fusion glycoprotein of
 RT Newcastle disease virus.";
 RL J. Gen. Virol. 67:2685-2694(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89204698; PubMed=2705298;
 RA Toyoda T., Sakaguchi T., Hirota H., Gotoh B., Kuma K., Miyata T.,
 RA Nagai Y.;
 RT "Newcastle disease virus evolution. II. Lack of gene recombination in
 RT generating virulent and avirulent strains.";
 RL Virology 169:273-282(1989).
 CC -1- FUNCTION: THIS PROTEIN DIRECTS FUSION OF VIRAL AND CELLULAR

MEMBRANES.
CC -!- SUBUNIT: THE MATURE FORM IS A DIMER OF POLYPEPTIDES F-1 AND F-2
CC LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE PARAMYXOVIRUSES FUSION GLYCOPROTEIN
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X04719; CAA28426.1; -
DR EMBL: M24697; AAA46648.1; -
DR EMBL: A03663; CAA00288.1; -
DR PIR: A27008; VGNZNY.
DR PIR: F46329; F46329.
DR HSSP: P04849; 1SVF.
DR InterPro: IPR000776; Fusion_gly.
DR Pfam: PF00523; Fusion_gly; 1.
KW Glycoprotein; Fusion protein; Transmembrane; Envelope protein; Signal;
KW Lipoprotein; Palmitate.
FT SIGNAL 1 25
FT CHAIN 26 553 FUSION GLYCOPROTEIN F0.
FT CHAIN 26 116 F2 PROTEIN.
FT CHAIN 117 553 F1 PROTEIN.
FT TRANSMEM 117 136 POTENTIAL.
FT DOMAIN 137 500 EXTRACELLULAR.
FT TRANSMEM 501 527 POTENTIAL.
FT DOMAIN 527 553 CYTOPLASMIC.
FT CARBOHYD 85 85 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 366 366 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 447 447 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT LIPID 523 523 PALMITATE (POTENTIAL).
SQ SEQUENCE 553 AA; 59041 MW; 63BFD1692AFDE191 CRC64;

Query Match 24.4%; Score 55.5; DB 1; Length 553;
Best Local Similarity 33.3%; Pred. No. 15;
Matches 16; Conservative 7; Mismatches 20; Indels 5; Gaps 1;

QY 4 TSLKNDP-----KEIKTDELEIIIGSGSLSTFFRLNRSFTQALGK 46
DB 433 TRLSGEPDVTYQKNISIDQSVIITGNLDISTELGNVNSISNALNK 480

RESULT 7
VGLF_NDVH4 STANDARD; PRT; 553 AA.
ID VGLF_NDVH4
AC P33613;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fusion glycoprotein precursor [contains: Fusion glycoprotein F2;
F. Fusion glycoprotein F1].
GN Newcastle disease virus (strain BI-Hitchner/47) (NDV).
OS Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Rubulaviruses.
OX NCBI_TaxID=11181;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204898; PubMed=2705298;
RA Toyoda T., Sakaguchi T., Hirota H., Gotoh B., Kuma K., Miyata T.,
Nagai Y.;
RT "Newcastle disease virus evolution. II. Lack of gene recombination in
RT generating virulent and avirulent strains.";
RL Virology 169:273-282(1989).
CC -!- FUNCTION: THIS PROTEIN DIRECTS FUSION OF VIRAL AND CELLULAR

MEMBRANES.
CC -!- SUBUNIT: THE MATURE FORM IS A DIMER OF POLYPEPTIDES F-1 AND F-2
CC LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE PARAMYXOVIRUSES FUSION GLYCOPROTEIN
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M24695; AAA46646.1; -
DR PIR: D46329; D46329.
DR HSSP: P04849; 1SVF.
DR InterPro: IPR000776; Fusion_gly.
DR Pfam: PF00523; Fusion_gly; 1.
KW Glycoprotein; Fusion protein; Transmembrane; Envelope protein; Signal;
KW Lipoprotein; Palmitate.
FT SIGNAL 1 25
FT CHAIN 26 553 FUSION GLYCOPROTEIN F0.
FT CHAIN 26 116 F2 PROTEIN.
FT CHAIN 117 553 F1 PROTEIN.
FT TRANSMEM 117 136 POTENTIAL.
FT DOMAIN 137 500 EXTRACELLULAR.
FT TRANSMEM 501 527 POTENTIAL.
FT DOMAIN 527 553 CYTOPLASMIC.
FT CARBOHYD 85 85 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 366 366 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 447 447 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT LIPID 523 523 PALMITATE (POTENTIAL).
SQ SEQUENCE 553 AA; 58909 MW; F163B9B/CFPB4493 CRC64;

Query Match 24.4%; Score 55.5; DB 1; Length 553;
Best Local Similarity 33.3%; Pred. No. 15;
Matches 16; Conservative 7; Mismatches 20; Indels 5; Gaps 1;

QY 4 TSLKNDP-----KEIKTDELEIIIGSGSLSTFFRLNRSFTQALGK 46
DB 433 TRLSGEPDVTYQKNISIDQSVIITGNLDISTELGNVNSISNALNK 480

RESULT 8
VGLF_NDVL STANDARD; PRT; 553 AA.
ID VGLF_NDVL
AC P33614;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fusion glycoprotein precursor [contains: Fusion glycoprotein F2;
F. Fusion glycoprotein F1].
GN Newcastle disease virus (strain Las/46) (NDV).
OS Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Rubulaviruses.
OX NCBI_TaxID=11184;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204898; PubMed=2705298;
RA Toyoda T., Sakaguchi T., Hirota H., Gotoh B., Kuma K., Miyata T.,
Nagai Y.;
RT "Newcastle disease virus evolution. II. Lack of gene recombination in
RT generating virulent and avirulent strains.";
RL Virology 169:273-282(1989).
CC -!- FUNCTION: THIS PROTEIN DIRECTS FUSION OF VIRAL AND CELLULAR
CC MEMBRANES.
CC -!- SUBUNIT: THE MATURE FORM IS A DIMER OF POLYPEPTIDES F-1 AND F-2
CC LINKED BY A DISULFIDE BOND.

CC -1- SIMILARITY: BELONGS TO THE PARAMYXOVIRUSES FUSION GLYCOPROTEIN
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M24696; AAA46647.1; -
 DR PIR: E46329; E46329.
 DR HSSP: P04849; 1SVF.
 DR InterPro: IPR000776; Fusion_gly.
 DR Pfam: PF00523; Fusion_gly; 1.
 KW Glycoprotein; Fusion protein; Transmembrane; Envelope protein; Signal;
 KW Lipoprotein; Palmitate.
 FT SIGNAL 1 25
 FT CHAIN 26 553
 FT CHAIN 26 116
 FT CHAIN 117 553
 FT TRANSMEM 117 136
 FT DOMAIN 137 500
 FT TRANSEM 501 527
 FT DOMAIN 528 553
 FT CARBOHYD 85 85
 FT CARBOHYD 191 191
 FT CARBOHYD 366 366
 FT CARBOHYD 447 447
 FT CARBOHYD 471 471
 FT LIPID 523 523
 SO SEQUENCE 553 AA; 59047 MW; 04E113853B91DF CAC64;
 Query Match 24.4%; Score 55.5; DB 1; Length 553;
 Best Local Similarity 33.3%; Pred. No. 15;
 Matches 16; Conservative 7; Mismatches 20; Indels 5; Gaps 1;
 Db 433 TLRSGEFDVYTKNISTQDSQVITGNLDISTELGWNNSISALNK 480
 Y 4 TSLKNDP-----KEIKTDELEIIIGSGSLSTFPRFNRSTQALGK 46
 D 433 TLRSGEFDVYTKNISTQDSQVITGNLDISTELGWNNSISALNK 480
 RESULT 9
 VGLF_NDVTG STANDARD; PRT; 553 AA.
 AC P12571;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Fusion glycoprotein precursor. [Contains: Fusion glycoprotein F2;
 DE Fusion glycoprotein F1].
 GN F.
 OS Newcastle disease virus (strain Texas 9.b./48) (NDV).
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
 OC Paramyxoviridae; Paramyxovirinae; Rubulavirus.
 OX NCBI_TaxID=11189;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86265873; PubMed=3388773;
 RA Schaper U.M., Fuller F.J., Ward M.D.W., Mehrotra Y., Stone H.O.,
 RA Stipp B.R., de Buyscher E.V.;
 RT Nucleotide sequence of the envelope protein genes of a highly
 RT virulent, neurotropic strain of Newcastle disease virus.";
 RL Virology 165:291-295(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89204898; PubMed=2705298;
 RA Toyoda T., Sakaguchi T., Hirota H., Gotoh B., Kuma K., Miyata T.,
 RA Nagai Y.;
 RT "Newcastle disease virus evolution. II. Lack of gene recombination in
 RT generating virulent and avirulent strains.";

RL Virology 169:273-282(1989).
 CC -1- FUNCTION: THIS PROTEIN DIRECTS FUSION OF VIRAL AND CELLULAR
 CC MEMBRANES.
 CC -1- SUBUNIT: THE MATURE FORM IS A DIMER OF POLYPEPTIDES F-1 AND F-2
 CC LINKED BY A DISULFIDE BOND.
 CC -1- SIMILARITY: BELONGS TO THE PARAMYXOVIRUSES FUSION GLYCOPROTEIN
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M23407; AAA46642.1; -
 DR EMBL: M24698; AAA46649.1; -
 DR PIR: B29201; VGNZGB.
 DR PIR: G46329; G46329.
 DR HSSP: P04849; 1SVF.
 DR InterPro: IPR000776; Fusion_gly.
 DR Pfam: PF00523; Fusion_gly; 1.
 KW Glycoprotein; Fusion protein; Transmembrane; Envelope protein; Signal;
 KW Lipoprotein; Palmitate.
 FT SIGNAL 1 25
 FT CHAIN 26 553
 FT CHAIN 26 116
 FT CHAIN 117 553
 FT TRANSEM 117 136
 FT DOMAIN 137 500
 FT TRANSEM 501 527
 FT DOMAIN 528 553
 FT CARBOHYD 85 85
 FT CARBOHYD 191 191
 FT CARBOHYD 366 366
 FT CARBOHYD 447 447
 FT CARBOHYD 471 471
 FT LIPID 523 523
 SO SEQUENCE 553 AA; 59017 MW; 5F537269F6P9B9 CRC64;
 Query Match 24.4%; Score 55.5; DB 1; Length 553;
 Best Local Similarity 33.3%; Pred. No. 15;
 Matches 16; Conservative 7; Mismatches 20; Indels 5; Gaps 1;
 Db 433 TLRSGEFDVYTKNISTQDSQVITGNLDISTELGWNNSISALNK 480
 Y 4 TSLKNDP-----KEIKTDELEIIIGSGSLSTFPRFNRSTQALGK 46
 D 433 TLRSGEFDVYTKNISTQDSQVITGNLDISTELGWNNSISALNK 480
 RESULT 10
 CB32_YEAST
 ID CB32_YEAST STANDARD; PRT; 608 AA.
 AC P40969;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Centromere DNA-binding protein complex CBP3 subunit B (Centromere
 DE protein 3).
 GN CBP3 OR CEP3 OR YMR168C OR YW8520.17C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetia; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=AC 502;
 RX MEDLINE=95045413; PubMed=7957085;
 RA Lechner J.;
 RA "A zinc finger protein, essential for chromosome segregation,
 RA constitutes a putative DNA binding subunit of the Saccharomyces
 RT cerevisiae kinetochore complex, Cbf3.";

RL EMBO J. 13:5203-5211(1994).
 RN (2)
 RP SEQUENCE FROM N.A.
 RX MEDLINE-95181532; PubMed-7876302;
 RA Strunilov A.V., Kingsbury J., Koshland D.;
 RT "CBF3 encodes a centromere protein of *Saccharomyces cerevisiae*";
 RL J. Cell Biol. 128:749-760(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972;
 RA Hunt S., Bowman S., Barrell B.G., Rejandream M.A.;
 RL Submitted (May-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: ESSENTIAL FOR CHROMOSOME SEGREGATION AND MOVEMENT OF
 CC CENTROMERES ALONG MICROTUBULES. IT PLAYS A ROLE IN THE ATTACHMENT
 CC OF CHROMOSOMES TO THE SPINDLE. CBF3 BINDS SELECTIVELY TO A HIGHLY
 CC CONSERVED DNA SEQUENCE CALLED CDEIII, FOUND IN CENTROMERES AND IN
 CC SEVERAL PROMOTERS.
 CC -1- SUBUNIT: CBF3 IS FORMED OF FOUR SUBUNITS, CBF3A (CBF2), CBF3B
 CC (CBF3), CBF3C (CTF13) AND CBF3D.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- SIMILARITY: CONTAINS 1 ZN(2)-CYS(6), FUNGAL-TYPE BINDING
 CC CLUSTER DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X81396; CAA57159.1; -
 DR EMBL: U12339; AAA57074.1; -
 DR EMBL: Z49705; CAA89804.1; -
 DR SGD: S0004778; CBF3.
 DR InterPro: IPR001138; ZN2_Cy6_fungal.
 DR Pfam: PF00172; ZN2_Cy6; 1.
 DR SMART: SM00066; GAL4; 1.
 DR PROSITE: PS00463; ZN2_Cy6_FUNGAL_1; 1.
 DR PROSITE: PS50048; ZN2_Cy6_FUNGAL_2; 1.
 KW Nuclear protein; Chromosomal protein; DNA-binding; Zinc;
 KW Metal-binding; Centromere.
 FT DNA BIND 14 42 ZN(2)-CYS(6); FUNGAL-TYPE.
 SQ SEQUENCE 608 AA; 71358 MW; 2E24A0508080A09B CRC64;

 Query Match 24.28; Score 55; DB 1; Length 608;
 Best Local Similarity 39.58; Pred. No. 19;
 Matches 15; Conservative 6; Mismatches 15; Indels 2; Gaps 1;

 QY 7 LKNDKFEIKTDELEIIGSGSISTFPLFRFSFTQ 42
 DB 541 LQNDKRIELKNDKEMFSLIKGSLVPLNKRQESILE 578

 RESULT 11
 ID HIS8_METUA STANDARD; PRT; 373 AA.
 AC 058365;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable histidinol-phosphate aminotransferase (EC 2.6.1.9) (imidazole
 DE acetyl-phosphate transaminase).
 GN HIS8 OR M00935.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
 OC Methanococcus.
 OX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE-96337999; PubMed-8686087;

RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Belch C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, *Methanococcus*
 RT *jannaschii*";
 RL Science 273:1058-1073(1996).
 CC -1- CATALYTIC ACTIVITY: L-histidinol-phosphate + 2-oxoglutarate = 3-
 CC (imidazol-4-yl)-2-oxopropyl phosphate + L-glutamate (By
 CC SIMILARITY).
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
 CC -1- PATHWAY: EIGHTH STEP IN HISTIDINE BIOSYNTHETIC PATHWAY.
 CC -1- SIMILARITY: BELONGS TO CLASS-II OF PYRIDOXAL-PHOSPHATE-DEPENDENT
 CC AMINOTRANSFERASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U67539; AAB98960.1; -
 DR TIGR: M00953; -
 DR InterPro: IPR001511; Aminotran_1.
 DR InterPro: IPR001917; Aminotran_2.
 DR Pfam: PF00155; aminotran_1_2; 1.
 DR PROSITE: PS00599; AA_TRANSF_CLASS_2; FALSE_NEG.
 KW Histidine biosynthesis; Transferase; Aminotransferase;
 KW Pyridoxal phosphate; Complete proteome.
 FT BINDING 231 231 PYRIDOXAL PHOSPHATE (PROBABLE).
 SQ SEQUENCE 373 AA; 42961 MW; 369DB5EB4C671218 CRC64;

 Query Match 24.08; Score 54.5; DB 1; Length 373;
 Best Local Similarity 39.48; Pred. No. 13;
 Matches 13; Conservative 5; Mismatches 12; Indels 3; Gaps 1;

 QY 7 LKNDKFEIKTDELEIIGSGS---LSTFPLRF 36
 DB 79 MKELSKFLVNDENITVGGGDADELIDITIRTF 111

 RESULT 12
 ID DNAA_RICCN STANDARD; PRT; 463 AA.
 AC 092856;
 DT 01-MAR-2002 (Rel. 41, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Chromosomal replication initiator protein dnaa.
 GN DNAA OR RC0916.
 OS Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsia conorii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsia.
 OX NCBI_TaxID=781;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Ma11sh 7;
 RX MEDLINE-21442074; PubMed-11557893;
 RA Ogata H., Audic S., Renseto-Audiffren P., Fournier P.-E., Barbe V.,
 RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
 RA Raoult D.;
 RT "Mechanisms of evolution in *Rickettsia conorii* and *R. prowazekii*";
 RL Science 293:2093-2098(2001).
 CC -1- FUNCTION: Plays an important role in the initiation and regulation
 CC of chromosomal replication. Binds to the origin of replication. It
 CC binds specifically double-stranded DNA at a 9 bp consensus (dnaa

CC box): 5'-TATAC(C/A)(C/A)-3'. Dnaa binds to ATP and to acidic
 CC phospholipids (by similarity).
 CC -1- SIMILARITY: BELONGS TO THE DNAA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE008645; AL03454.1;
 CC InterPro: IPR001957; Bac_Dnaa.
 CC Pfam: PF00308; bac_dnaa.1.
 CC PROSITE: PS01008; DNAA.1.
 CC DNA replication; DNA-binding; ATP-binding; Complete proteome.
 CC NP BIND 168 175 ATP (POTENTIAL).
 CC FT SEQUENCE 463 AA; 52942 MW; B48D2FEA3A70A8EA CRC64;
 CC
 CC Query Match 24.0%; Score 54.5; DB 1; Length 463;
 CC Best Local Similarity 28.2%; Pred. No. 16;
 CC Matches 11; Conservative 8; Mismatches 17; Indels 3; Gaps 1;
 CC
 CC 2 KRTSLKNDKFEIKT---DELEIIIGSGSLSTFERLFN 37
 CC Db 213 KEVISFEKERSVDYLMIDIFGCKDSTOEFEFFHTFN 251
 CC
 CC RESULT 13
 CC SYFB_PORPU STANDARD; PRT; 720 AA.
 CC ID PS1346;
 CC AC 01-OCT-1996 (Rel. 34, Created)
 CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Phenylalanyl-tRNA synthetase beta chain (EC 6.1.1.20) (Phenylalanine--
 CC tRNA ligase beta chain) (phers).
 CC GN PHET.
 CC OS Porphyra purpurea.
 CC OC Chloroplast.
 CC OX Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
 CC NCBI_TaxID=2787;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=AVONPORT;
 CC RA Reith M.E., Munnholland J.;
 CC RT "Complete nucleotide sequence of the Porphyra purpurea chloroplast
 CC genome.";
 CC RL Plant Mol. Biol. Rep. 13:333-335(1995).
 CC CC -1- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +
 CC dihydrophosphate + L-phenylalanyl-tRNA(Phe).
 CC CC -1- SUBUNIT: Tetramer of two alpha and two beta chains (by
 CC similarity).
 CC CC -1- SUBCELLULAR LOCATION: Chloroplast.
 CC CC -1- SIMILARITY: BELONGS TO THE PHENYLALANYL-TRNA SYNTHETASE BETA
 CC CHAIN FAMILY. SUBFAMILY 1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: 038804; AAC0232.1;
 CC KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 CC Chloroplast.
 CC SO SEQUENCE 720 AA; 82592 MW; 495EA9947117F62A CRC64;

CC Query Match 24.0%; Score 54.5; DB 1; Length 720;
 CC Best Local Similarity 31.5%; Pred. No. 26;
 CC Matches 17; Conservative 7; Mismatches 9; Indels 21; Gaps 2;
 CC
 CC 6 SLKNDKFEIKTDELEIIIGSGSLSTFERLFN 39
 CC Db 495 NLKRN-KIETTKLAIILGSLIRSEMSPAHSLMWYKARKGIENFFKLNKS 547
 CC
 CC RESULT 14
 CC CXB2_MOUSE STANDARD; PRT; 226 AA.
 CC ID CXB2_MOUSE
 CC AC 000977;
 CC DT 01-APR-1993 (Rel. 25, Created)
 CC DT 01-APR-1993 (Rel. 25, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Gap junction beta-2 protein (Connexin 26) (Cx26).
 CC GN GJB2 OR CXN-26.
 CC OS Mus musculus (Mouse).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC NCBI_TaxID=10090;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE=92354576; PubMed=1322820;
 CC RA Willecke K., Nicholson B.J., Dahl E., Kozjek G., Hennemann H.;
 CC RT "Molecular cloning of mouse connexins26 and -32: similar genomic
 CC organization but distinct promoter sequences of two gap junction
 CC genes.";
 CC RL Eur. J. Cell Biol. 58:81-89(1992).
 CC RN [2]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE=91285228; PubMed=2060697;
 CC RA Nishi M., Kumar N.M., Gilula N.B.;
 CC RT "Developmental regulation of gap junction gene expression during
 CC mouse embryonic development.";
 CC RL Dev. Biol. 146:117-130(1991).
 CC CC -1- FUNCTION: ONE GAP JUNCTION CONSISTS OF A CLUSTER OF CLOSELY PACKED
 CC PAIRS OF TRANSMEMBRANE CHANNELS, THE CONNEXONS, THROUGH WHICH
 CC MATERIALS OF LOW MW DIFFUSE FROM ONE CELL TO A NEIGHBORING CELL.
 CC CC -1- SUBUNIT: A CONNEXON IS COMPOSED OF A HEXAMER OF CONNEXINS.
 CC CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC CC -1- TISSUE SPECIFICITY: LIVER, KIDNEY, INTESTINE, LUNG, SPLEEN,
 CC STOMACH, TESTIS, & BRAIN, BUT NOT HEART & ADULT SKELETAL MUSCLE.
 CC CC -1- SIMILARITY: BELONGS TO THE CONNEXIN FAMILY. BETA-TYPE (GROUP 1)
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M81445; AAA37495.1;
 CC EMBL: M63803; AAA37276.1;
 CC PIR: C49769; C49769.
 CC MGD: MGI:95720; Gjb2.
 CC DR InterPro: IPR000500; Connexin.
 CC DR Pfam: PF00029; connexin.1.
 CC DR PRINTS: PR00206; CONNEXIN.
 CC DR SMART: SM00037; CNX.1.
 CC DR PROSITE: PS00407; CONNEXINS_1; 1.
 CC DR PROSITE: PS00408; CONNEXINS_2; 1.
 CC KW Gap junction; Transmembrane.
 CC FT DOMAIN 1 19
 CC FT DOMAIN 2 20
 CC FT TRANSMEM 40
 CC FT DOMAIN 41 75
 CC FT TRANSMEM 76 96
 CC FT DOMAIN 97 143
 CC FT TRANSMEM 144 164
 CC FT DOMAIN 165 189
 CC EXTRACELLULAR (POTENTIAL).
 CC POTENTIAL.
 CC EXTRACELLULAR (POTENTIAL).
 CC POTENTIAL.
 CC CYTOPLASMIC (POTENTIAL).
 CC POTENTIAL.
 CC CYTOPLASMIC (POTENTIAL).
 CC POTENTIAL.
 CC EXTRACELLULAR (POTENTIAL).

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:14:14 ; Search time 45.14 Seconds

(without alignments)
176.291 Million cell updates/sec

Title: US-09-833-017-2

Sequence: 1 MKKTLKNDKFKETDELE.....GSLSTFFRLNRSFTQALGK 46

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	227	100.0	46	2	099Q15
2	220	96.9	46	2	09APK7
3	209	92.1	43	2	09APK5
4	69	30.4	850	16	097N40
5	63.5	28.0	648	2	09APK6
6	59.5	26.2	275	16	09CPK8
7	58	25.6	48	2	033596
8	57.5	25.3	1009	16	099WU3
9	57.5	25.3	453	3	09C1A3
10	57.5	25.3	516	10	09SUK0
11	57.5	25.3	516	10	09LXK3
12	57	25.1	591	2	09AHY8
13	56.5	24.9	407	2	09R006
14	56.5	24.9	407	2	09P331
15	56.5	24.9	408	16	09K1F7
16	56.5	24.9	593	16	092G06

17	56	24.7	1250	16	09PPL7	09PPL7 campylobact
18	55.5	24.4	125	12	041368	041368 newcastle d
19	55.5	24.4	125	12	041472	041472 newcastle d
20	55.5	24.4	146	16	092EBO	092EBO listeria in
21	55.5	24.4	553	12	090339	090339 newcastle d
22	55.5	24.4	553	12	09DLD4	09DLD4 newcastle d
23	55.5	24.4	553	12	09MMH7	09MMH7 newcastle d
24	55.5	24.4	553	12	092150	092150 newcastle d
25	55.5	24.4	553	12	09WLE2	09WLE2 newcastle d
26	55.5	24.4	553	12	09WLE1	09WLE1 newcastle d
27	55.5	24.4	553	12	091HX4	091HX4 newcastle d
28	55.5	24.4	553	12	091AH8	091AH8 newcastle d
29	55.5	24.4	837	10	09M819	09M819 arabidopsis
30	55.5	24.4	1304	5	09U0H4	09U0H4 plasmodium
31	55	24.2	48	2	033615	033615 streptococc
32	54.5	24.0	230	16	09CGX0	09CGX0 streptococc
33	54.5	24.0	255	2	09AKR3	09AKR3 rickettsia
34	54.5	24.0	255	2	09AKR6	09AKR6 rickettsia
35	54.5	24.0	398	16	09ALN0	09ALN0 streptococc
36	54.5	24.0	463	16	092H56	092H56 rickettsia
37	54.5	24.0	819	5	097330	097330 plasmodium
38	54	23.8	48	2	033651	033651 streptococc
39	54	23.8	235	2	09X929	09X929 vibrio para
40	54	23.8	243	16	097M22	097M22 clostridium
41	54	23.8	2353	2	P71401	P71401 haemophilus
42	53.5	23.6	125	12	041386	041386 newcastle d
43	53.5	23.6	125	12	041388	041388 newcastle d
44	53.5	23.6	165	1	000392	000392 methanococc
45	53.5	23.6	176	2	09RJ74	09RJ74 streptomyc

ALIGNMENTS

RESULT	ID	099Q15	PRELIMINARY:	PRT:	46 AA.
AC	099Q15	099Q15			
DT	01-JUN-2001	(TREMBLrel. 17, Created)			
DT	01-JUN-2001	(TREMBLrel. 17, Last sequence update)			
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)			
DE	COMPETENCE STIMULATING PROTEIN PRECURSOR.				
GN	COMC.				
OS	Streptococcus mutans.				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;				
OX	NCBI_TaxID=1309;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-GH14, H7, LT11, NG8, AND UA159;				
RX	MEDLINE-21142515; PubMed-11208787;				
RA	Li Y.H., Lau P.C.Y., Lee J.H., Ellen R.P., Vitkovitch D.G.;				
RT	"Natural Genetic Transformation of Streptococcus mutans Growing in				
RT	Biofilms."				
RL	J. Bacteriol. 183:897-908(2001).				
DR	EMBL; AF277152; AAK01542.1; -				
DR	EMBL; AF277153; AAK01543.1; -				
DR	EMBL; AF277155; AAK01545.1; -				
DR	EMBL; AF277156; AAK01546.1; -				
DR	EMBL; AF277157; AAK01547.1; -				
DR	Interpro: IPR004288; COMC.				
DR	Plan. PF03047; COMC; 1.				
FT	CHAIN 26				
SQ	SEQUENCE 46 AA; 5211 MW; 38FA62B6F78FC3BF CRC64;				

Query Match 100.0%; Score 227; DB 2; Length 46;
Best Local Similarity 100.0%; Pred No. 5.3e-22;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKKTLKNDKFKETDELEIIIGSGSLSTFFRLNRSFTQALGK 46
DB 1 MKKTLKNDKFKETDELEIIIGSGSLSTFFRLNRSFTQALGK 46

```

RESULT 2
ID 09APK7 PRELIMINARY; PRT; 46 AA.
AC 09APK7;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
GN COMP.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BM71;
RX MEDLINE=21142515; PubMed=11208787;
RA Li Y.H., Lau P.C.Y., Lee J.H., Ellen R.P., Cvitekovich D.G.;
RT "Natural Genetic Transformation of Streptococcus mutans Growing in
RT Biofilms.";
RL J. Bacteriol. 183:897-906(2001).
DR EMBL: AF277151; AAK01541.1;
FT CHAIN 26 46
SQ SEQUENCE 46 AA; 5195 MW; 38E0B9DB5B8FC3BF CRC64;

```

```

Query Match 96.8%; Score 220; DB 2; Length 46;
Best Local Similarity 97.8%; Pred. No. 4.2e-21;
Matches 45; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 MKKTLKNDKFKETDELEITIGSGSLSTFFRLNRSFTQALGK 46
DB 1 MKKTLKNDKFKETDELEITIGSGSLSTFFRLNRSFTQALGK 46

```

```

RESULT 3
ID 09APK6 PRELIMINARY; PRT; 43 AA.
AC 09APK6;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE COMPETENCE STIMULATING PROTEIN.
GN COMC.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JH1005;
RX MEDLINE=21142515; PubMed=11208787;
RA Li Y.H., Lau P.C.Y., Lee J.H., Ellen R.P., Cvitekovich D.G.;
RT "Natural Genetic Transformation of Streptococcus mutans Growing in
RT Biofilms.";
RL J. Bacteriol. 183:897-906(2001).
DR EMBL: AF277154; AAK01544.1;
FT CHAIN 26 43
SQ SEQUENCE 43 AA; 4927 MW; E6A78FC3BF6156C7 CRC64;

```

```

Query Match 92.1%; Score 209; DB 2; Length 43;
Best Local Similarity 97.7%; Pred. No. 1e-19;
Matches 42; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MKKTLKNDKFKETDELEITIGSGSLSTFFRLNRSFTQ 43
DB 1 MKKTLKNDKFKETDELEITIGSGSLSTFFRLNRSFTQ 43

```

```

RESULT 4

```

```

097N40
ID 097N40 PRELIMINARY; PRT; 850 AA.
AC 097N40;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE ABC TRANSPORTER, PERMEASE PROTEIN, PUTATIVE.
GN SP2231.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Durkin A.S., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Umayam L.A., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA Holtzapple E., Khourel H., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:498-506(2001).
DR EMBL: AE007510; AAK06279.1;
DR TIGR: SP2231;
KW Complete proteome.
SQ SEQUENCE 850 AA; 97303 MW; 1ADED613F06B5115 CRC64;

```

```

Query Match 30.4%; Score 69; DB 16; Length 850;
Best Local Similarity 33.3%; Pred. No. 2.5;
Matches 19; Conservative 10; Mismatches 16; Indels 12; Gaps 2;

```

```

QY 2 KKTLSLKNDKFKETDELEITIGSGSLSTFFRLNRSFTQALGK 46
DB 692 KKTFTDNVSEFNLGYTKRKFTNNVSPGNSQVSFESPTFYRIDTFTTEALQK 748

```

```

RESULT 5
ID 09L651 PRELIMINARY; PRT; 648 AA.
AC 09L651;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE HYPOTHETICAL LACTOCOCCIN 972 IMMUNITY PROTEIN (LACTIS PLASMID PBL1 DNA
DE FOR LACTOCOCCIN 972 OPERON).
GN LCB.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OG Plasmid PBL1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IPLA 972; PLASMID=PBL1;
RA Sanchez C., Hernandez de Rojas A., Martinez B., Arqueles M.E.,
RA Suarez J.E., Rodriguez A., Mayo B.;
RT "Nucleotide sequence of PBL1, a bacteriocin-producing plasmid from
RT Lactococcus lactis IPLA 972.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=IPLA 972; PLASMID=PBL1;
RA Suarez J.E.;
RT Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=IPLA 972; PLASMID=PBL1;

```

RX MEDLINE=20055640; PubMed-10589723;
 RA Martinez B., Fernandez M., Rodriguez A., Suarez J.E.;
 RT "Synthesis of lactococcin 972, a bacteriocin produced by *Lactococcus*
 RT *lactis* IPLA 972, depends on the expression of a plasmid-encoded
 RT bicistronic operon."
 RL Microbiology 145:3155-3161(1999).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IPLA 972; PLASMID=PB1;
 RA Martinez B.;
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF242367; AAF64055.1;
 DR EMBL: AJ002203; CAC03468.1;
 KW Plasmid.
 SQ SEQUENCE 648 AA; 74139 MW; 79530E501D6371F7 CRC64;

Query Match 28.0%; Score 63.5; DB 2; Length 648;
 Best Local Similarity 46.2%; Pred. No. 9.6;
 Matches 18; Conservative 8; Mismatches 10; Indels 3; Gaps 2;

QY 5 LSKNDFEIKTDELEIIIGSGSL--STF-FRLFRSF 40
 DB 528 LSVKNRFEIKSQQLFIVTSSIALISSTFLFLYLNKITY 566

RESULT 6
 Q9CPE8 PRELIMINARY; PRT; 275 AA.
 AC Q9CPE8;
 DT 01-JUN-2001 (TREMUREL. 17, Created)
 DT 01-JUN-2001 (TREMUREL. 17, Last sequence update)
 DT 01-DEC-2001 (TREMUREL. 19, Last annotation update)
 DE PROC.
 GN PROC OR PM0095
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Pasteurella.
 OX NCBI_TaxID=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PM70;
 RX MEDLINE=21145866; PubMed-11248100;
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
 RT "Complete genomic sequence of *Pasteurella multocida* PM70."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
 DR EMBL: AE006044; AAK02179.1;
 DR InterPro: IPR000304; P5CR.
 DR Pfam: PF01089; P5CR; 1.
 DR PROSITE: PS00521; P5CR; UNKNOWN_1.
 KW Complete proteome.
 SQ SEQUENCE 275 AA; 29556 MW; 011C7F11AD31A7D7 CRC64;

Query Match 26.2%; Score 59.5; DB 16; Length 275;
 Best Local Similarity 29.6%; Pred. No. 12;
 Matches 16; Conservative 7; Mismatches 16; Indels 15; Gaps 1;

QY 6 LSKNDFEIKTDELEIIIGSGSLSTFRLFRNSFTQAL 44
 DB 141 LKREYKQFQDLNNAVGKTCWTOADNHTITAGSGSPATFLFLENMGOAL 194

RESULT 7
 Q33596 PRELIMINARY; PRT; 48 AA.
 AC Q33596;
 DT 01-JAN-1998 (TREMUREL. 05, Created)
 DT 01-JAN-1998 (TREMUREL. 05, Last sequence update)
 DT 01-DEC-2001 (TREMUREL. 19, Last annotation update)
 DE COMPETENCE STIMULATING PEPTIDE PRECURSOR.
 GN COMC.
 OS Streptococcus anginosus.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus;
 OX NCBI_TaxID=1328;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCIC 10713;
 RX MEDLINE=98012953; PubMed-9352904;
 RA Havarstein L.S., Hakenbeck R., Gaustad P.;
 RT "Natural competence in the genus *Streptococcus*: Evidence that
 RT streptococci can change phenotype by interspecies recombinational
 RT exchanges."
 RL J. Bacteriol. 179:6589-6594(1997).
 DR EMBL: AJ000864; CAA04341.1;
 DR InterPro: IPR004288; COMC.
 DR Pfam: PF03047; COMC; 1.
 KW Signal.
 FT SIGNAL 1 32
 FT CHAIN 33 48
 SQ SEQUENCE 48 AA; 5621 MW; EB6DF7AEDE4E94DB CRC64;

Query Match 25.6%; Score 58; DB 2; Length 48;
 Best Local Similarity 30.2%; Pred. No. 2.8;
 Matches 16; Conservative 9; Mismatches 16; Indels 12; Gaps 2;

QY 1 MKKTLKLN-----DKREIKTDELEIIIGSGSLSTFRLFRNSFTQALG 46
 DB 1 MKKLFRAKKEVKAVERKELNDEQLDKITIGDSRIRMGF----DKSLGK 48

RESULT 8
 Q99WU3 PRELIMINARY; PRT; 1009 AA.
 AC Q99WU3;
 DT 01-JUN-2001 (TREMUREL. 17, Created)
 DT 01-JUN-2001 (TREMUREL. 17, Last sequence update)
 DT 01-DEC-2001 (TREMUREL. 19, Last annotation update)
 DE SA0272 PROTEIN (HYPOTHETICAL PROTEIN SAV0283).
 GN SA0272 OR SAV0283.
 OS *Staphylococcus aureus* (strain N315), and
 OS *Staphylococcus aureus* (strain Mu50).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Staphylococcus.
 OX NCBI_TaxID=158879, 158878;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES-S. aureus (strain N315), and S. aureus (strain Mu50);
 RX MEDLINE=21311952; PubMed-11418146;
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
 RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-O., Ito T.,
 RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
 RA Mizutani M., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
 RA Sekimizu K., Hiraoka H., Kuhara S., Goto S., Yanuzaki J.,
 RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
 RA Hattori M., Ogatawara N., Hayashi H., Hiratsuka K.;
 RT "Whole genome sequencing of methicillin-resistant *Staphylococcus*
 RT *aureus*."
 RL Lancet 357:1225-1240(2001).
 DR EMBL: AP003130; BAB41496.1;
 DR EMBL: AP003358; BAB56445.1;
 KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 1009 AA; 114781 MW; 16AD06D170CA31FE CRC64;

Query Match 25.6%; Score 58; DB 16; Length 1009;
 Best Local Similarity 34.3%; Pred. No. 79;
 Matches 12; Conservative 8; Mismatches 15; Indels 0; Gaps 0;

QY 9 NDKREIKTDELEIIIGSGSLSTFRLFRNSFTQAL 43
 DB 202 NDKPELFTDTLVNSISKNDIKRWQTYNKSLSLA 236

	RESULT	9	
ID	O9CIA3		PRT:
AC	O9CIA3	PRELIMINARY;	453 AA.
DT	01-JUN-2001 (TREMBLrel. 17,	(Created)	
DT	01-JUN-2001 (TREMBLrel. 17,	Last sequence update)	
DT	01-OCT-2001 (TREMBLrel. 18,	Last annotation update)	
DE	SULFATE TRANSPORTER SULZ-LA	(FRAGMENT).	
OS	Saccharomyces pastorianus (yeast).		
OC	Eukaryota; Fungi; Ascomycota;	Saccharomycotina; Saccharomycetes;	
CC	Saccharomycetales; Saccharomycetaceae;	Saccharomycetes.	
OX	NCBI_TaxID=27292;		
RN	[1]		
RA	JAMES FROM N.A.		
RP	Mrs A.B., Slaughter C., Meaden P.G.;		
RT	"Characterization of partial coding sulfate transporter sequences from		
RL	Saccharomyces pastorianus and Saccharomyces bayanus."		
DR	Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.		
DR	EMBL; AF364410; AAAK32879.1;	-	
DR	Interpro: IPR001902; Sulfate_transp.		
FT	Non_TER	1	
FT	NON_TER	453	
SO	SEQUENCE	453 AA; 49519 MW; 137117AD0D12BD31F CRC64;	
	Query Match	25.3%; Score 57.5; DB 3; Length 453;	
	Best local Similarity	36.7%; Pred.No. 38;	
	Matches	18; Conservative 6; Mismatches 18; Indels 7; Gaps 2;	
OY	1 MKTFLSKNDKEIKTELEIIGGSGSLSPFFRR-----FNRSFTAQ 43 : : : : : : : : 227 ISSFGKVANDKYVPPDEL-TALGVSNLTGFNNAYPATGSFSNSALK 274		
	RESULT	10	
ID	O9SUKO	PRELIMINARY;	PRT: 516 AA.
AC	O9SUKO;		
DT	01-MAY-2000 (TREMBLrel. 13,	Created)	
DT	01-MAY-2000 (TREMBLrel. 13,	Last annotation update)	
DT	01-DIC-2001 (TREMBLrel. 19,	Last annotation update)	
DE	GLUCOSE-6-PHOSPATE 1-DEHYDROGENASE (EC 1.1.1.45) (GGPD).		
G6P	AraBiodopsis thaliana (Mouse-ear cross). Arabidopsidis thailiana (Mouse-ear cross).		
OS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatocypita; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OC	eucosids II; Brassicales; Brasiaticaceae; Arabidopdis.		
OX	NCBI_TaxId=3702;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=g9364543; PubMed=10437832;		
RA	Wendt U.K.; Henschl R.; Lange C.; Pietersma M.; Wenderoth I., von Schewen A.;		
RT	"Evidence for functional convergence of redox regulation in GGPDH isoforms of cyanobacteria and higher plants.";		
RL	Plant Mol. Biol. 40:487-494(1999).		
-1-	CATALYTIC ACTIVITY: D-GULOSE 6-PHOSPATE + NAD(+)= D-GLUONO- DELTA-LACTONE 6-PHOSPATE + NADH.		
CC	-1- PATHWAY: FIRST STEP IN Pentose Phosphate Pathway.		
CC	-1- SIMILARITY: BELONGS TO THE GLUCOSE-6-PHOSPATE DEHYDROGENASE FAMILY.		
CC	EMBL: AJ010970; CAB52674.1; -		
DR	HSSP: PIL111; IDGP.		
DR	InterPro: IPR001282; GGPD.		
DR	pfam: PF00479; GGPD; 1.		
DR	pfam: PF02781; GGPD.C; 1.		
DR	PRINTS: PR00079; GGPDHGRNAS.		
DR	ProDom: PD001129; GGPD; 1.		
KM	PROSITE: PS00069; GFP_DEHYDOENASE; 1.		
GO	Glucose metabolism; NADP; Oxidoreductase.		
GO	Sequence	516 AA; AV623780FP053CALC CRC64;	

```

Query Match: 25.3%; Score 57.5; DB 10; Length 516;
Best Local Similarity 41.5%; Pred. No. 44;
Matches 22; Conservative 7; Mismatches 11; Indels 13; Gaps 5

OY 1 MKKTLISKND--FRE---IKTDELEII--GGSGSL--TF---FRLFNRSF 40
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8 MEKRSSTLKNDSFVKEYNPTETGSGSLIIVLGAAGDLAKKTFPALFNLFHQGF 60

RESULT 11
O9LK23
ID 09LK23 PRELIMINARY; PRT; 516 AA.
AC 09LK23
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2001 (TREMBLrel. 15, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE GLUCOSE-6-PHOSPHATE 1-DEHYDROGENASE (EC 1.1.1.49) (56PD).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots; Rosidae;
OC eucots II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxID=3702;
[1]
RN RP
RE SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RA Kaneo T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RA Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN RP
RE SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RA MEDLINE-20363099; PubMed-10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety p1,
RT TAC and BAC clones.";
RT DNA Res. 7:217-221(2000).
CC -1- CATALYTIC ACTIVITY: D-GLUCOSE 6-PHOSPHATE + NADP(+) = D-GLUCO-
CC -1- DELTA-LACTONE 6-PHOSPHATE + NADPH.
CC -1- PATHWAY: FIRST STEP IN PENTOSE PHOSPHATE PATHWAY.
CC -1- SIMILARITY: BELONGS TO THE GLUCOSE-6-PHOSPHATE DEHYDROGENASE
CC FAMILY.
DR EMBL: AP000381; BAB02125.1; -.
DR HSSP; P11411; 1DRC.
DR InterPro; IPR001282; G6PD.
DR Pfam; PF00479; G6PD; 1.
DR Pfam; PF02781; G6PD_C; 1.
DR PRINTS; PR00079; G6PDHGRGNASE.
DR PRODOM; PD001129; G6PD; 1.
DR PROSITE; PS00069; G6PD_DEHYDROGENASE; 1.
KW Glucose metabolism; NADP; Oxidoreductase.
SQ SEQUENCE 516 AA; 59157 MW; 407E4259EFC20E2 CRC64;

Query Match 25.3%; Score 57.5; DB 10; Length 516;
Best Local Similarity 41.5%; Pred. No. 44;
Matches 22; Conservative 7; Mismatches 11; Indels 13; Gaps 5.

OY 1 MKKTLISKND--FRE---IKTDELEII--GGSGSL--TF---FRLFNRSF 40
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8 MEKRSSTLKNDSFVKEYNPTETGSGSLIIVLGAAGDLAKKTFPALFNLFHQGF 60

RESULT 12
O9AHY9
ID 09AHY9 PRELIMINARY; PRT; 591 AA.
AC 09AHY9;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE YFAA (FRAGMENT).
OS Photorhabdus luminescens (Xenorhabdus luminescens).

```

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OX Photobacterium
 RN NCBI_TaxID=29488;
 RP SEQUENCE FROM N.A.
 RC STRAIN-NC19, ATCC29304;
 RX MEDLINE-21225535; PubMed-11325940;
 RA Cliche T.A., Brinton S.B., Horswill A.R., Engen J.C.;
 RT "A Phosphopantetheinyl Transferase Homolog Is Essential for
 Photobacterium luminescens to Support Growth and Reproduction of the
 Endosymbiotic Nematode Heterorhabdus bacteriophora.";
 RL J. Bacteriol. 183:3117-3126(2001).
 DE EMBL; AF288085; AAK16098.1;
 GN HSPB; P39435; IKAS.
 OS Interpro: IPR000794; Ketoacyl-synt.
 DR Pfam: PF00109; ketoacyl-synt_C.1
 DR PROSITE: PS00606; B_KETOACYL_SINTHASE; UNKNOWN_1.
 FT NON_TER 591 591
 SQ SEQUENCE 591 AA; 63878 MW; AFI451C6DD500891 CRC64;

Query Match 25.1%; Score 57; DB 2; Length 591;
 Best Local Similarity 34.3%; Pred. No. 59;
 Matches 12; Conservative 7; Mismatches 16; Indels 0; Gaps 0;

OY 12 KEIKTDELEIIGSGSLSTFFRLFNRSFTQALGK 46
 DB 149 QRVNRDRIGVILGGGNGMTLFLSLAGROOTPYLRK 183

RESULT 13
 O9R006 PRELIMINARY; PRT; 407 AA.
 AC Q9R006; 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE BUSA.
 GN BUSA.
 OS Lactococcus lactis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OX Lactococcus
 RN NCBI_TaxID=1358;
 RP SEQUENCE FROM N.A.
 RC STRAIN-NCDO 763;
 RX MEDLINE-99445450; PubMed-10515910;
 RA Obis D., Guillot A., Girpon J.C., Renault P., Bolotin A., Mistou M.Y.;
 RT "Genetic and biochemical characterization of a high-affinity betaine
 uptake system (BUSA) in Lactococcus lactis reveals a new functional
 organization within bacterial ABC transporters.";
 RL J. Bacteriol. 181:6238-6246(1999).
 DE EMBL; AF139575; AAF04258.1;
 GN EMBL; AF139575; AAF04258.1;
 OS Interpro: IPR003593; AAA.
 DR Interpro: IPR003439; ABC_transportr.
 DR Interpro: IPR001687; ATP_GTP_A.
 DR Interpro: IPR000644; CBS.
 DR Pfam: PF00005; ABC_tran; 1.
 DR SMART; SM00382; AAA; 1.
 DR SMART; SM00116; CBS; 2.
 DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
 KW ATP-binding; Transport.
 SQ SEQUENCE 407 AA; 45696 MW; D5DFBC0F8764B3A9 CRC64;

Query Match 24.9%; Score 56.5; DB 2; Length 407;
 Best Local Similarity 45.7%; Pred. No. 46;
 Matches 16; Conservative 3; Mismatches 15; Indels 1; Gaps 1;

OY 13 EIKTDELEIIGSGSLSTFFRLFNRSFTQALGK 46
 DB 149 QRVNRDRIGVILGGGNGMTLFLSLAGROOTPYLRK 183

DB 50 EINEGEIIVIMLGSCKSTLLRLNRLIEPTSGK 84

RESULT 14
 O9L931 PRELIMINARY; PRT; 407 AA.
 AC O9L931; 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE OPVUA.
 GN OPVUA.
 OS Lactococcus lactis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OX Lactococcus
 RN NCBI_TaxID=1358;
 RP SEQUENCE FROM N.A.
 RC STRAIN-MG1363;
 RX MEDLINE-20393175; PubMed-10939245;
 RA Bouvier J., Bordes P., Romeo Y., Fourcans A., Bouvier I.,
 RT "Characterization of OpvA, a glycine-betaine uptake system of
 Lactococcus lactis.";
 RL J. Mol. Microbiol. Biotechnol. 2:199-205(2000).
 DE EMBL; AF184955; AAF0473.1;
 GN Interpro: IPR003593; AAA.
 DR Interpro: IPR003439; ABC_transportr.
 DR Interpro: IPR001687; ATP_GTP_A.
 DR Pfam: PF00005; ABC_tran; 1.
 DR SMART; SM00382; AAA; 1.
 DR SMART; SM00116; CBS; 2.
 DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
 KW ATP-binding; Transport.
 SQ SEQUENCE 407 AA; 45653 MW; 656D9097AE58407F CRC64;

Query Match 24.9%; Score 56.5; DB 2; Length 407;
 Best Local Similarity 45.7%; Pred. No. 46;
 Matches 16; Conservative 3; Mismatches 15; Indels 1; Gaps 1;

OY 13 EIKTDELEIIGSGSLSTFFRLFNRSFTQALGK 46
 DB 50 EINEGEIIVIMLGSCKSTLLRLNRLIEPTSGK 84

RESULT 15
 O9K1F7 PRELIMINARY; PRT; 408 AA.
 AC O9K1F7; 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 18, Last annotation update)
 DE OPVUA (BETAINE ABC TRANSPORTER ATP BINDING PROTEIN).
 GN OPVUA OR BUSA.
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis)
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OX Lactococcus
 RN NCBI_TaxID=1360;
 RP SEQUENCE FROM N.A.
 RC MEDLINE-20318987; PubMed-10860977;
 RA van der Heide T., Poolman B.;
 RT "Osmoregulated ABC-transport system of Lactococcus lactis senses water
 stress via changes in the physical state of the membrane.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:7102-7106(2000).
 DE EMBL; AF1403;
 GN STRAIN-1L1403;
 OS MEDLINE-21235186; PubMed-11337471;
 KW Bolotin A., Wincker P., Manger S., Jailion O., Malarne K.,

RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
lactis ssp. lactis IL1403.";
RL Genome Res. 11:731-753(2001).
CC -) SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC (ABC TRANSPORTERS).
DR EMBL: AF234619; AAF37878.1; -;
DR EMBL: AE006375; AK05550.1; -;
DR InterPro: IPR003593; AAA.
DR InterPro: IPR003439; ABC_transport.
DR InterPro: IPR001687; ATP_GTP_A.
DR InterPro: IPR000644; CBS.
DR Pfam: PF00005; ABC_tran; 1.
DR Pfam: PF00571; CBS; 2.
DR SMART: SM00382; AAA; 1.
DR SMART: SM00116; CBS; 2.
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Complete proteome; Transport.
SQ SEQUENCE 408 AA; 45750 MW; EBA8A87BE563B88 CRC64;

Query Match 24.9%; Score 56.5; DB 16; Length 408;
Best Local Similarity 45.7%; Pred. No. 46;
Matches 16; Conservative 3; Mismatches 15; Indels 1; Gaps 1;
QY 13 EIKTDELEITIGGSGS-LSTFRLFNRSFTQALGK 46
DB 50 EINEGEIFVIMGLSGSKSTLLRLNRLLEPTSGK 84

Search completed: July 30, 2002, 15:20:30
Job time: 376 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:14:11 ; Search time 53.42 Seconds

(without alignments)
916.951 Million cell updates/sec

Title: US-09-833-017-4

Sequence: 2173

1 MNEALMILSLGLTYLTLVLF.....KTSLSQSNHNLFKQLLIK 441

Scoring table:

BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched:

747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters:

747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq.032802:*

- 1: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1980.DAT:*
- 2: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1981.DAT:*
- 3: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1982.DAT:*
- 4: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1983.DAT:*
- 5: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1984.DAT:*
- 6: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1985.DAT:*
- 7: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1986.DAT:*
- 8: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1987.DAT:*
- 9: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1988.DAT:*
- 10: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1989.DAT:*
- 11: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1990.DAT:*
- 12: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1991.DAT:*
- 13: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1992.DAT:*
- 14: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1993.DAT:*
- 15: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1994.DAT:*
- 16: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1995.DAT:*
- 17: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1996.DAT:*
- 18: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1997.DAT:*
- 19: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1998.DAT:*
- 20: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1999.DAT:*
- 21: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA2000.DAT:*
- 22: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	752	34.6	446	20	AAV14065
2	752	34.6	446	20	AAV03650
3	752	34.6	446	20	AAW89437
4	752	34.6	446	20	AAW89434
5	398	18.3	175	19	AAV85911
6	378.5	17.4	432	22	AAW81843
7	258	11.9	108	19	AAW36531
8	189	8.8	106	19	AAW36530
9	189	8.7	106	18	AAV11170
10	151	6.9	543	22	AAW98391
11	145.5	6.7	442	20	AAW83369

12	145.5	6.7	442	21	AAV57664	Streptococcus pneu
13	143	6.6	861	19	AAW69977	Staphylococcus aur
14	143	6.6	909	22	AAU37397	Staphylococcus aur
15	142.5	6.6	363	22	AAU37208	Staphylococcus aur
16	141.5	6.5	363	19	AAV5788	Staphylococcus aur
17	139	6.4	605	22	ABW52567	Escherichia coli p
18	135.5	6.2	478	21	AAV95040	Escherichia coli p
19	135.5	6.2	1120	22	AAU15085	Candida albicans p
20	134	6.2	565	19	AAW89798	Protein encoded by
21	134	6.2	565	19	AAW71554	H. pylori GHP 128
22	133	6.2	565	22	AAW6337	Helicobacter poly
23	133	6.1	298	22	AAW62102	H. pylori HPS024 p
24	125.5	5.8	4134	20	AAV31946	S. epidermidis ope
25	124.5	5.7	952	22	ABG29521	Plasmodium falcipa
26	124	5.7	451	22	AAW81558	Novel human diagno
27	122	5.6	436	22	AAU03143	S. epidermidis ope
28	119	5.5	600	17	AAW95269	Streptococcus pyog
29	119	5.5	610	20	AAV06667	Nisin nist gene pr
30	117	5.4	618	21	AAW28685	Nisin C of lactoba
31	116	5.3	1477	20	AAV19981	Staphylococcus war
32	116	5.3	1494	20	AAV19980	B. burgdorferi ant
33	115.5	5.3	1079	22	ABW69644	B. burgdorferi ant
34	115	5.3	850	19	AAW57445	Drosophila melano
35	115	5.3	1167	22	ABW8194	A. thermophilum th
36	114.5	5.3	904	20	AAW6351	Drosophila melano
37	114	5.2	457	20	AAV05736	Human DNA toll-11
38	114	5.2	543	22	ABW47270	Staphylococcus aure
39	113.5	5.2	443	21	AAW81506	Streptococcus faec
40	113.5	5.2	451	20	AAW28601	Streptococcus pneu
41	113	5.2	1865	17	AAW03515	Histidine kinase p
42	112.5	5.2	471	20	AAV20061	Human DOKK180 prot
43	112.5	5.2	490	20	AAV20060	B. burgdorferi ant
44	112.5	5.2	1346	21	AAW18236	B. burgdorferi ant
45	112	5.2	794	22	AAW0426	Plasmodium falcipa

ALIGNMENTS

RESULT 1

ID	AAV14065	standard; Protein: 446 AA.
XX	AAV14065;	
XX	16-JUL-1999 (first entry)	
DE	S. pneumoniae Ornithine carbamoyltransferase.	
XX		
KW	Ornithine carbamoyltransferase; argf/arcb family polypeptide; diagnosis;	
KW	Infection; therapy; immune response; otitis media; conjunctivitis;	
KW	pneumonia; bacteraemia; sinusitis; pleural empyema; endocarditis;	
KW	meningitis; matrix protein adhesion.	
OS	Streptococcus pneumoniae.	
XX		
PN	EP913476-A2.	
XX	06-MAY-1999.	
PD		
XX	22-OCT-1998; 98EP-0203571.	
PF		
XX	30-OCT-1997; 97US-0961536.	
PR		
XX	(SMK) SMITHKLINE BEECHAM CORP.	
PA		
XX	Brown JR, Zalacain M;	
XX		
PT	WPI: 1999-256630/22.	
DR	N-PSDB: AAX57966.	
XX		
PT	New Streptococcus pneumoniae ornithine carbamoyltransferase (argf/arcb) polypeptide and polynucleotide; us	

CC This sentence is the streptococcus pneumoniae ornithine
CC carbamoyltransferase protein (I) of the invention. The ornithine
CC carbamoyltransferase protein is a (argF/arcB) family polypeptide. (I)
CC and polynucleotides (II) encoding it are useful for diagnosing diseases
CC due to an infection of an organism with the ornithine
CC carbamoyltransferase gene by determining the nucleic acid sequence
CC encoding (I), and/or analysing for the presence or amount of (I). They
CC can diagnose the stage and type of infection. (I) is also useful for
CC screening for compounds which affect activity of the protein by measuring
CC the binding to (I) and observing the stimulation or inhibition of the
CC polypeptide function. These can be used in treatment to inhibit or
CC enhance ornithine carbamoyltransferase activity. In addition to direct
CC administration of ornithine carbamoyltransferase polypeptides to treat
CC conditions associated with a lack of ornithine carbamoyltransferase
CC polypeptide, or direct administration of antisense sequences to prevent
CC expression, (I) and antibodies against it induce an immune response to
CC immunise and prevent disease. Anti-(I) antibodies induced by the
CC polypeptide are also useful for isolating clones expressing (I), or for
CC purifying the polypeptide by affinity chromatography. Diseases diagnosed
CC prevented or treated include otitis media, conjunctivitis, pneumonia,
CC bacteraemia, sinusitis, pleural empyema, endocarditis and especially
CC meningitis. (I), (II) and their (ant)agonists can prevent adhesion of
CC bacteria to matrix proteins, and are useful for use on wounds and body
CC implants to prevent bacterial infection.
CC Note: the specification states that this sequence is identical to
CC that shown in AYL14067.

Query Match	34.6%;	Score 752;	DB 20;	Length 446;
Best Local Similarity	39.2%;	Pred. No. 3e-58;		
Matches 172;	Conservative	86;	Mismatches 169;	Indels 12;
				Gaps 5

```

QY 12 LFWLFWL--FLFLFLSKXSNVTLKKELFLFSLNFMILMAVT-----MVNVNLEPAE 64
Db 6 lllvnlvnlkvlvlffkngvgigltfdrflkaflklflgliftftcfglavskyslyfe 65
QY 65 PLVFLALSIYLNKONSLSINIEYGLLPVASSDLERALLFFILDTG-QGIV--MGSSITT 12
Db 66 plflgltgslfllllrglplkklallfgyllfpmllvelfyrgvsyflprflsgvgdgdnprfl 12
QY 122 TYMLERFAGLNLSTLFLSPVNDVIGLKKOSLKKMKVKKRLLPMMNTMLKLYLLOVLYVE 18
Db 126 lllmflvcflvI--vflkwldydcflrflrfltdbfgfxskllkhamagayllvmgslsyfe 18
QY 182 SYNVYLPRLKFRKKFVVVUYLLFLFLILFSLSOYTKOKOVONEIMAKOEAKOINNTIOYSQOE 24
Db 184 yeglgigtvtrhlllvfyllffmgslkklldyrlkexlqeelngqvltrydmearyshie 24
QY 242 SLKVDKDSFRADYDNLNITSRLGIGIEKMDKLASIEKTYHQILEKNGHLOOTPRNIGHLANI 30
Db 244 elykeitsfndbyhnltslrlrglfeedeomegikelydsavlrsdsqkldhkydlgrlvi 30
QY 302 ONDVAVKGLSLKLTLEONKKNIVANVEVSKTOLPEMELFTLTSLCDNAIEAFESL 36
Db 304 rdlralkslgkflkarekrlvtrleovpeelvgemslldtlvlslldcnaievsaaes 36
QY 362 NPEIOLAEFKKNGSIYFIQNTSTKEKODVSKIFKFNSTGKSNRGIGLAKVNHILEHPR 42
Db 364 qphvsiaflkngqetfliensikeegldlseifsgasskeergyulvlyvmkiveshr 42
QY 422 KTSLOTSNNHHLFKOOLLII 440
Db 424 ntnlntscqrvltqvylv 442

```

RESULT

DT 09-JUN-1999 (first entry)

DE S. pneumoniae histidine kinase polypeptide.

..... Histidine kinase; two component signal transduction system; TCSTS;
KW

KW bacteremia; sinusitis; pleural empyema; endocarditis; meningitis;

XX
XX

DIRECTOR, FBI

Streptococcus pneumoniae.

EP892059-A2.
PN
XX

PD 20-JAN-1999.
XX

PF 17-JUN-1998; 98EP-0304782.
YY

PR 20-JUN-1997; 97US-0879941.

PA (SMIK) SMITHKLINE BEECHAM CORP.

XX

XX	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
----	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----

DR N-P\$DB; AAX29247.

aa
PT New Histidine kinase polypeptide and polynucleotide - useful as

PT Streptococcus pneumoniae infections, especially meningitis

XX
XX
PS
Claim 6: Page 9: 37pp; English.

This represents a histidine kinase (HK) polypeptide which is a component of the two component signal transduction system (TCSS) in bacteria. The gene encoding the mature polypeptide can be isolated from *S. pneumoniae* 0109933. Host cells containing a vector comprising the nucleic acid are used for the recombinant expression of the protein. HK polypeptides and polynucleotides are useful for diagnosing diseases related to over or underexpression of HK protein by identifying mutations in the HK gene, or determining HK polypeptide or mRNA expression levels due to an infection of an organism with the HK gene. The polypeptides are also useful for screening for compounds which affect activity of the protein. These can be used in treatment to inhibit (antagonist i.e. antibacterial drugs) or enhance (agonist) HK activity. The polypeptides (administered directly, in a vector and as a vaccine) and antibodies induce an immune response to immunise and prevent disease. Diseases diagnosed, prevented or treated include: bacterial infections, especially *S. pneumoniae* infections, which cause otitis media, conjunctivitis, bacteraemia, sinusitis, pleural empyema, endocarditis and especially meningitis. The polypeptides, polynucleotides and their (ant)agonists can prevent adhesion of bacteria to matrix proteins, and are useful for use on wounds and body implants to prevent bacterial infection.

Sequence 446 AA:

Query Match	34.68;	Score 752;	DB 20;	Length 446;
Best Local Similarity	39.28;	Pred. No. 3e-58;		
Matches 172; Conservative	86;	Mismatches 169;	Indels 12;	Gaps 5,

QY 12 LLLVTVL-FLLEFSKSNVTLSKRETLFEESNLIINIAV-----MNVNLFFPAE 64
Db 6 llyvalvnglkvivtfknsgnlgltfdcfidafklkellglltfttcqflavskylysfiie 65

QY 65 PLVTAISITLNQNQLSNIITFGCLLPVASDGLFRRAILIFELDTG-OGIY-MGSSITT 124H
Db 66 plfgtgstflllpglpkrlklllyglfpmulveltyrsvsfyfpplfgldvgdggnpfl 125H

QY 122 TYMIEFAGIALSYFLSVENVDIGRLKSLTKMKVKRLPMNITMLLYLLIOVLVIE 181
 CC the binding, or stimulation or inhibition of RR protein activity.
 CC Treatment of diseases related to protein activity can be treated by
 CC administering antagonists to prevent expression, and RR proteins to
 CC enhance expression levels of the protein. RR proteins and polynucleotides
 CC can be administered as a vaccine to protect against disease, and can
 CC prevent adhesion of bacteria to matrix proteins, and are useful for use
 CC on wounds and body implants to prevent bacterial infection. Antibodies
 CC are useful for treating bacterial infections, and can isolate or identify
 CC clones expressing RR proteins, and can purify the proteins by affinity
 CC chromatography. Diseases diagnosed, prevented or treated include: otitis
 CC media; conjunctivitis; pneumonia; bacteremia; sinusitis; pleural empyema;
 CC endocarditis and particularly meningitis, and infections caused by
 CC Streptococcus pneumoniae.
 CC XX
 SQ Sequence 446 AA:
 QY 362 NPEIOLAFKKNGSIVFIIONSTKEKOIDVSKIFKENVSTKSGNSRGIGLAKVNHILEHYP 421
 Db 364 qphvsiaflknagqetfllensikeegidseifsgasskgeergvgylyvmkiveshp 423
 QY 422 KTSLOTSNNHHLEFKOLLII 440
 Db 424 ntlnltcgnqvtirgyltv 442
 RESULT 3
 AAM89437
 ID AAM89437 standard; Protein; 446 AA.
 AC AAM89437;
 XX
 DT 22-MAR-1999 (first entry)
 DE Streptococcus pneumoniae histidine kinase.
 KW Streptococcus pneumoniae; response regulator; histidine kinase;
 KW infection; antibacterial; otitis media; conjunctivitis; pneumonia;
 KW bacteremia; sinusitis; pleural empyema; endocarditis; meningitis.
 OS Streptococcus pneumoniae.
 XX
 PN EP885901-A2.
 PD 23-DEC-1998.
 PF 09-JUN-1998; 98EP-0304562.
 PR 13-JUN-1997; 97US-0874138.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PA (SMIK) SMITHKLINE BEECHAM PLC.
 PI Wallis NG;
 DR WPI, 1999-037018/04.
 DR N-PSDB; AAV82066.
 XX
 PT New response regulator polypeptide and polynucleotide - useful as
 PT diagnostic reagents and for prevention and treatment of
 PT Streptococcus pneumoniae infections
 XX
 PS Disclosure: Page 16; 40pp; English.
 CC The present sequence represents histidine kinase from Streptococcus
 CC pneumoniae, which is cognate of the response regulator (RR) of the
 CC present invention. RR proteins and polynucleotides are useful as
 CC research reagents and as materials for discovery of treatments of, and
 CC for diagnosing diseases, particularly for diagnosing the stage and type
 CC of an infection using PCR. Diseases related to expression or activity of
 CC RR proteins can be diagnosed by determining the nucleotide sequence
 CC encoding RR, or by analysing for the presence or amount of RR proteins.
 CC RR proteins can be used to screen for agonists and antagonists

CC (antibacterial compounds) which interact with RR proteins by observing
 CC the binding, or stimulation or inhibition of RR protein activity.
 CC Treatment of diseases related to protein activity can be treated by
 CC administering antagonists to prevent expression, and RR proteins to
 CC enhance expression levels of the protein. RR proteins and polynucleotides
 CC can be administered as a vaccine to protect against disease, and can
 CC prevent adhesion of bacteria to matrix proteins, and are useful for use
 CC on wounds and body implants to prevent bacterial infection. Antibodies
 CC are useful for treating bacterial infections, and can isolate or identify
 CC clones expressing RR proteins, and can purify the proteins by affinity
 CC chromatography. Diseases diagnosed, prevented or treated include: otitis
 CC media; conjunctivitis; pneumonia; bacteremia; sinusitis; pleural empyema;
 CC endocarditis and particularly meningitis, and infections caused by
 CC Streptococcus pneumoniae.
 CC XX
 SQ Sequence 446 AA:
 Query Match 34.6%; Score 752; DB 20; Length 446;
 Best Local Similarity 39.2%; Pred. No. 3e-58;
 Matches 172; Conservative 86; Mismatches 169; Indels 12; Gaps 5;
 QY 12 LLYLYMYL--FLFLFSKVSNTLSKKELETLFISNFIIMIAVY----MNVNVLFPYPAE 64
 Db 6 llylalynglkivifkvnqigltfdrifkafiklilgtftqflavskyslyfie 65
 QY 65 PLVEFIALSTLYLNKONSLSNIFGGLPVASSDLPFRRAIIFFLDGT--QGIV--MGSSITP 121
 Db 66 plfgjglsflllgrlpkklilfyglfpmllvelfyrgvsytlvflfgslvdggnbfl 125
 QY 122 TYMIEFAGIALSYFLSVENVDIGRLKSLTKMKVKRLPMNITMLLYLLIOVLVIE 181
 Db 126 lmfivfcivl--vflkwldydftrlrrefldtqfksltklmwagaylyvmgslyle 183
 QY 182 SYNVIPFLKFRKRVIVYLLIFLLISLSQYTKOKVONEMAKOEAQIRNITOYSQOIE 241
 Db 184 yegqisgtvtrnhllvfyllffmgikldtylkeklgeelngedtlrydmerysrhe 243
 QY 242 SLVKDISEFRHDYINLTLSRLGIENKDLASIEKTYHOLEKTHODPTRYNIGHLANI 301
 Db 244 elykeisfrhdytnltsrlglsdeedmegikelydsvlrdsqklgndkydlgrlvni 303
 QY 302 QNDVAKGILSAKILLEAKONKRIAVNVEVSSKIQLPMEMLDPTIISLIDNMLEAFESL 361
 Db 304 rdtalksllagkfkikareknlvfnvepveelqvegmllidflivslidcnalevsas 363
 QY 362 NPEIOLAFKKNGSIVFIIONSTKEKOIDVSKIFKENVSTKSGNSRGIGLAKVNHILEHYP 421
 Db 364 qphvsiaflknagqetfllensikeegidseifsgasskgeergvgylyvmkiveshp 423
 QY 422 KTSLOTSNNHHLEFKOLLII 440
 Db 424 ntlnltcgnqvtirgyltv 442
 RESULT 4
 AAM89434
 ID AAM89434 standard; Protein; 446 AA.
 AC AAM89434;
 XX
 DT 22-MAR-1999 (first entry)
 DE Streptococcus pneumoniae histidine kinase.
 KW Streptococcus pneumoniae; response regulator; histidine kinase;
 KW infection; antibacterial; otitis media; conjunctivitis; pneumonia;
 KW bacteremia; sinusitis; pleural empyema; endocarditis; meningitis.
 OS Streptococcus pneumoniae.
 XX
 PN EP885963-A2.

PD 23-DEC-1998.
 XX
 PF 17-JUN-1998; 98BP-0304797.
 XX
 PR 20-JUN-1997; 97US-0879528.
 XX
 PA (SMIR) SMITHKLINE BEECHAM CORP.
 PA (SMIR) SMITHKLINE BEECHAM PLC.
 XX
 PI Wallis NW;
 XX
 DR WPI: 1999-037058/04.
 DR N-PDB; AAV82059.
 XX
 PT New response regulator polypeptide and polynucleotide - useful as
 PT diagnostic reagents and for prevention and treatment of
 PT Streptococcus pneumoniae infections
 PS Disclosure; Page 15; 37pp; English.
 XX
 CC The present sequence represents histidine kinase from Streptococcus
 CC pneumoniae, which is cognate of the response regulator (RR) of the
 CC present invention. RR proteins and polynucleotides are useful as research
 CC reagents and materials for discovery of treatments of and diagnostics
 CC for disease, and particularly for diagnosing the stage and type of an
 CC infection by PCR. Diseases related to expression or activity of RR can
 CC be diagnosed by determining the nucleotide sequence encoding RR, or by
 CC analysing for the presence or amount of RR protein. RR proteins can be
 CC used to screen for agonists and antagonists which interact with RR
 CC proteins by observing the binding, or stimulation or inhibition of RR
 CC protein activity. The agonist can be used in treatment to inhibit RR
 CC activity, and RR proteins can be administered to treat conditions
 CC associated with a lack of RR protein. Diseases can be prevented by
 CC inoculating with RR protein, or a vector which expresses RR protein. RR
 CC proteins and polynucleotides can prevent adhesion of bacteria to matrix
 CC proteins, and are useful for use on wounds and body implants to prevent
 CC bacterial infection. RR antibodies are useful for treating bacterial
 CC infections, and can isolate or identify clones expressing RR proteins,
 CC and can purify the proteins by affinity chromatography. Diseases
 CC diagnosed, prevented or treated include: otitis media; conjunctivitis;
 CC pneumonia; bacteremia; sinusitis; pleural empyema; endocarditis and
 CC particularly meningitis, and infections caused by Streptococcus
 CC pneumoniae.
 CC
 CC Sequence 446 AA:
 SQ

Query Match 34.6%; Score 752; DB 20; Length 446;
 Best Local Similarity 39.2%; Pred. No. 3e-58;
 Matches 172; Conservative 86; Mismatches 169; Indels 12; Gaps 5;

12 LLYLVVL--FLLEFSKVSNTLSKRELLFSLNFMIAVY----WVNNLFPAE 64
 Db 6 LLYLVAVNGKLVIVIFKVNIGYIFDILIKAFILKILIGLITFTGFLAVSKYSYIFIE 65
 QY 65 PLVIALSLYLRKQNSLNTIFGGLPVASSDLFRRAIIFILDGT-QGIV-WGSSITT 121
 Db 66 PLFGIGSLSTLLRTPKILIIYGLFPMILVELYRGVSFVLPFGQGVADGQPIEL 125
 QY 122 TWIEFAGIALSYLFSVNVNDIGRLKMSLKKMKVRRKLPMNTMLYLLLOVLYVE 181
 Db 126 LMLTFCEIVL--VILKWLQYDFLTRITREIDTGFQSLKINWAGAYLVMSISLYIE 183
 QY 182 SYNVIFPLKFRKRVVIVLILFLISLSQYTRKQVONETMAQKQAINITQYSQOTE 241
 Db 184 YEGGISTEVRLHLLVIFLLFFMGIKRLDYIKELKEELNGEQLTRYLDMEYGRHIE 243
 QY 242 SLKQDRSRFRHYLNTITSLRGLENKDLASTEKIYQIIEKKGHOLODRYVIGILANI 301
 Db 244 ELKELSRFRHYLNTITSLRGLENKDLASTEKIYQIIEKKGHOLODRYVIGILANI 303
 QY 302 QNDVAVGILSAKILEAKQNKRAVNVVSSKIQLPDMELDDFTITLSDNALTEAAFESE 361
 Db 302 QNDVAVGILSAKILEAKQNKRAVNVVSSKIQLPDMELDDFTITLSDNALTEAAFESE 361

Db 304 rdaJksllagkfikareknivfwepeelqvegmllldflivslldnaalewsaas 363
 QY 362 NPEICLAFFKKGSIYFLLIONSTREKQIDVSKIFKENVSTKSGNRIGLAKVNHLEHP 421
 Db 364 qpvstafllkngagctfllensikeegldiselfsgasckgeergvlytvmkliveshp 423
 QY 422 KTSLOTSNHHHLEKOLLII 440
 Db 424 ntnlntcgvfyrgvltcv 442

RESULT 5
 AAY85911
 ID AAY85911 standard; Protein: 175 AA.
 XX
 AC AAY85911;
 XX
 DT 10-APR-2000 (first entry)
 XX
 DE S. pneumoniae derived protein #120.
 XX
 KM Treatment; prevention; disease; diagnosis; gene therapy; screening;
 KM bacterial; antimicrobial; antibiotic; pathogenesis; infection.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN W09806734-A1.
 XX
 PD 19-FEB-1998.
 XX
 PF 15-AUG-1997; 97WO-US14436.
 XX
 PR 16-AUG-1996; 96US-0024022.
 XX
 PA (SMIR) SMITHKLINE BEECHAM CORP.
 XX
 PI Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;
 PI Stodola RK;
 XX
 DR WPI: 1998-159452/14.
 DR N-PDB; AAY96278.
 XX
 PT Streptococcus pneumoniae proteins and related DNA - useful for
 PT screening compounds for antibacterial activity
 PS Claim 5; Page 404-405; 640pp; English.
 XX
 CC This invention describes novel isolated Streptococcus pneumoniae
 CC polynucleotides (see AA296173-296494) and their encoded proteins (see
 CC AA85792-Y86182). The DNA, vectors and host cells described in the
 CC method of the invention are useful for the recombinant expression of the
 CC polypeptides. The polypeptides are useful for treatment or prevention of
 CC disease, or diagnosis of disease related to expression or activity of
 CC such a polypeptide. They can also be used to screen for compounds which
 CC interact with and inhibit or activate such a polypeptide. The
 CC polypeptides (or DNA encoding them, via gene therapy) are also useful
 CC for inducing an immunological response in a mammal. The antagonists are
 CC useful to inhibit such bacterial polypeptides. The polypeptides are
 CC particularly useful to identify antimicrobial compounds and antibiotics.
 CC They are also useful to determine their role in pathogenesis of
 CC infection, dysfunction and disease.
 CC
 CC Sequence 175 AA:
 SQ

Query Match 18.3%; Score 398; DB 19; Length 175;
 Best Local Similarity 44.4%; Pred. No. 2.1e-27;
 Matches 76; Conservative 40; Mismatches 55; Indels 0; Gaps 0;

QY 270 LASEIKIYHOLEKRGHLODTRNIGHLANTQNDAYKGLSKILEAKQNKRAVNVSS 329
 Db 1 meqikeiyavldrsgskldqdkyolgivnlrdaJksllagkfikareknivfwepe 60

Novel Streptococcus pneumoniae proteins and related DNA - useful for diagnosing anti-microbial agents for treatment of bacterial

DR N-PSDB; AAX30770.
XX

[illegible]

CC	RESULT 10
AC	AAG98391
ID	AAG98391 standard; Protein; 543 AA.
XX	AAG98391;
AC	
XX	21-SEP-2001 (first entry)
DT	
XX	
DE	Escherichia coli protein sequence SEQ ID NO:439.
XX	
KW	Escherichia coli; identification; proliferation; microorganism;
KW	antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;
KW	bacterial growth inhibition.
XX	
OS	Escherichia coli.
XX	
PN	MO200148209-A2.
XX	
PD	05-JUL-2001.
XX	
PE	19-DEC-2000; 2000MC-US34419.
XX	
PR	23-DEC-1999; 99US-0173005.
XX	
PA	(ELIT-) ELITRA PHARM INC.
XX	
PI	Forsyth RA, Ohlsen KL, Zyskind JW;
XX	
DR	WPI; 2001-457376/49.
DR	N-PSDB; AAH81447.
XX	
PT	Novel nucleic acids encoding proteins required for Escherichia coli
PT	proliferation, useful for screening for antimicrobial agents -
XX	
PS	Claim 19; Page 558-559; 596pp; English.
XX	
CC	The present invention describes a purified or isolated nucleic acid
CC	sequence (I) consisting essentially of one of the 93 nucleotide sequences
CC	given in AAH81202 to AAH81294, where expression of the nucleic acid in a
CC	microorganism is capable of inhibiting proliferation of a microorganism.
CC	(I) have antibacterial and antibiotic activities, and can be used in
CC	gene therapy. Expression of (I) in a microorganism inhibits proliferation

[illegible]

RESULT	11
AAW83369	
ID	AAW83369 standard; Protein; 442 AA.
XX	
AC	AAW83369;
XX	
DT	17-FEB-1999 (first entry)
XX	
DE	Streptococcus pneumoniae histidine kinase.
XX	
KW	Streptococcus pneumoniae; histidine kinase; antibacterial; diagnosis;
KW	Streptococcal; Helicobacter pylori; infection; bacteriostatic; microbial;
KM	bacteriocidal; ulcer; gastritis; stomach cancer; gene therapy.
XX	
OS	Streptococcus pneumoniae.
XX	
PN	EP881286-A2.
XX	
PD	02-DEC-1998.
XX	
PF	26-MAY-1998; 98EP-030413B.
XX	

PR 30-MAY-1997; 97US-0048347.
 XX (SMIK) SMITHKLINE BEECHAM CORP.
 XX
 PI Biswas S, Throup J, Wallis NG, Zalacain M;
 XX
 DR WPI: 1999-001392/01.
 XX N-PSDB: AAV72649.
 XX

XX New Streptococcus pneumoniae Histidine kinase polypeptide and
 PT polynucleotide - useful as diagnostic reagents and for prevention
 PT and treatment of Streptococcal and Helicobacter pylori infections
 XX
 PS
 XX

Claim 1; Page 34-35; 44pp: English.

XX The present sequence represents histidine kinase (HK) isolated from
 CC Streptococcus pneumoniae. HK polynucleotides and polypeptides are useful
 CC for diagnosing susceptibility to diseases by detecting mutations or
 CC polymorphisms in the HK gene or analysing for the presence or amount
 CC of HK polypeptide expressed in a patient sample. HK PCR probes are
 CC useful for diagnosing diseases (especially Streptococcal), and can
 CC characterise the stage and the species or strain causing the infection.
 CC The HK probes can also determine the response of the infectious organism
 CC to drugs. HK polypeptides and polynucleotides are useful for screening
 CC for antagonists, agonists and drugs against infectious micro-organisms.
 CC HK agonists and antagonists are bacteriostatic and bacteriocidal
 CC compounds which can be used in treatment to enhance (agonist) or block
 CC (antagonist or antisense sequence) HK activity, therefore treating
 CC microbial (especially Streptococcal) diseases, ulcers and gastritis, and
 CC stomach cancer caused by Helicobacter pylori. Epitopes of HK polypeptides
 CC and polynucleotides are useful immunogens for producing anti-HK
 CC antibodies for vaccines to prevent bacterial infections, and HK
 CC polynucleotides can be used in genetic immunisation (gene therapy) to
 CC prevent infections. HK polypeptides and polynucleotides and their
 CC (ant)agonists can prevent adhesion of bacteria to matrix proteins, and
 CC are useful for use on wounds and body implants to prevent bacterial
 CC infection.
 XX
 SQ Sequence 442 AA;

Query Match 6.7%; Score 145.5; DB 20; Length 442;
 Best Local Similarity 21.7%; Pred. No. 0.00019;
 Matches 104; Conservative 70; Mismatches 154; Indels 151; Gaps 21;

QY 24 LFSKSNVTLTKKELTFSISNFMIMAVMWNLFYPAEPIYALSIYL-NRONSLS 82
 DB 6 lftkiflyfs-----lfsvlclhla-----yflfptyshrgctlg 46
 QY 83 LNIFFGLPVASDLFRRAIFFITD-----GTGIVMGSSITITVMEFFGIALSYLF 136
 DB 47 gkataiaaglegkd--rqslegvldysqtsdtkgvkge-----mte----- 87
 QY 137 LSVFNVDGRKDKSL-----TKMKVKRRLIPMNI--TMELYVLLI-----QVLY 178
 DB 88 -----dklevkdsldpdtqtslfierevktqdggtmlqflasmdlkseagisl 141
 QY 179 VIESNVNPTLKFRKPVYVYLLIFLLISLSQYKQKQVEMAKQAQR----- 231
 DB 142 gflpylliaflislalvayartivapllekrvttr-----tmddlsqyrlrvdskd 195
 QY 232 NITQYQQIESLYKD-----IRSFRRHYDNIITSLRLGIE 266
 DB 196 efgnklqeglnlyghlltvladlhekneallglekmkveflgashelktpasikille 255
 QY 267 N-----KD-----LASIKYHQLE-----KTGHQODPYVNIQHLANTONAV 306
 DB 256 nmremigrtykrdqlyvalyvelnhvlyqlslsvsqvelrdretldllqmtgnlvk 315
 QY 307 KGLISAKLLEAQ-----NKRIVANVEVSSKIQLPMEMLDFTTILSDCNAIAEARES 360
 DB 316 dyallakereidqansltbqaylmpsv-----mkl-----llslnlsnaikhsy-- 360

QY 361 LNPETQLAFKKNGSIVFIIONSTFKKODIVSKIFKENYSTGSGNIGLAVNHLEH 419
 DB 361 --pgglvtrigerel-fienccssegeqktaqtsidnasirvkxgsmgllfvksilleh 416

RESULT 12
 AAY57684
 ID AAY57684 standard; Protein: 442 AA.
 XX

AC AAY57684;

DT 16-MAR-2000 (first entry)

DE Streptococcus pneumoniae sensor histidine kinase.

KW Antibiotic; growth inhibition; bacterial autolysis; regulation;
 KW antibacterial; infection; inflammation.

OS Streptococcus pneumoniae.

PN W09957281-A2.

PD 11-NOV-1999.

PF 06-MAY-1999; 99WO-US09792.

PR 06-MAY-1998; 98US-0073541.

PA (SUJD-) ST JUDE CHILDREN'S RES HOSPITAL.

PI Novak R, Tuomanen EI;

DR WPI: 2000-062151/05.

DR N-PSDB: AA248094.

PT New nucleic acid and peptides, useful as antibiotic peptides

PS Claim 38; Page 128-130; 151pp: English.

XX The present invention describes peptide antibiotics which inhibit the
 CC growth of pneumococci, ABC transporter and two-component signal
 CC transduction system proteins from Streptococcus pneumoniae. The peptides
 CC are useful (especially in the form of a pharmaceutical composition) for
 CC the treatment of a bacterial infection or inflammation. Methods from the
 CC present invention are also useful for the identification of agents or
 CC drugs which are useful in preventing bacterial proliferation or kill
 CC bacterial cells e.g. potential antibiotics. The peptides and methods are
 CC also useful for providing important epidemiological tools. The peptide
 CC are especially useful for the prevention of any disease caused by a
 CC bacterium e.g. Streptococcus aureus, Acinetobacter, Enterococcus
 CC faecalis, Escherichia coli, Pseudomonas aeruginosa all of which can
 CC cause blood poisoning among other ailments. Mycobacterium tuberculosis
 CC and Neisseria gonorrhoeae which causes gonorrhoea. The peptides are also
 CC useful in the treatment of infections due to Streptococcus pneumoniae a
 CC bacterial species that causes blood poisoning, middle ear infections,
 CC pneumonia and meningitis in humans. The present sequence represents a
 CC sensor histidine kinase from the present invention.

SQ Sequence 442 AA;

Query Match 6.7%; Score 145.5; DB 21; Length 442;
 Best Local Similarity 21.7%; Pred. No. 0.00019;
 Matches 104; Conservative 70; Mismatches 154; Indels 151; Gaps 21;

QY 24 LFSKSNVTLTKKELTFSISNFMIMAVMWNLFYPAEPIYALSIYL-NRONSLS 82
 DB 6 lftkiflyfs-----lfsvlclhla-----yflfptyshrgctlg 46
 QY 83 LNIFFGLPVASDLFRRAIFFITD-----GTGIVMGSSITITVMEFFGIALSYLF 136
 DB 47 gkataiaaglegkd--rqslegvldysqtsdtkgvkge-----mte----- 87

```

QY 137 LSVFNVDIGRLKDSL-----TKMKYKKRLIPMNI--TMLVLYLI-----OVLY 178
      |||||
Db 88 -----dklevdsldpdrdrqtslfiieerevktqdgumtqflasmdtqkeaqsl 141
      |||||
QY 179 VIESVNVIPLEKFRKRVIVVILFLILISFLSQYTKQKQVONEIMAKQAIR-----231
      |||||
Db 142 qflpyelllasflslvaylvarclvapllleikvtrr-----mmdtdsgvrlrvqskd 195
      |||||
QY 232 NITQVSQOIESLYKD-----IRSPRHOYLNLTSLRGITE 266
      |||||
Db 196 efgnlkeqslsyqnlvliadlhekneallqlekmkveflrgssheklrpsakllile 255
      |||||
QY 267 N-----KD-----LASIEKTYHQILE---KTGHOLDPTRYNIGHLANIQNDAY 306
      |||||
Db 256 nmrenlgrkddqylgvalglvdelnhvllslssvgelrdrredldlqmtqnlvk 315
      |||||
QY 307 KGLISAKLIEAQ-----NKRIAVNVEVSSKIQLPPEMLDPTITLSTLCNATEAPES 360
      |||||
Db 316 dyallakereidqdnslthqeylnpsv-----mkl-----llsnllsnaikhsv-- 360
      |||||
QY 361 LNPETQLAEFKKNGSIVFIIONSTKEKQIDVSKIFKENVSTKGSNRGIGLAKVNHILEH 419
      |||||
Db 361 --pgjvirigeregel-fienescseegeklqafsfdsnarkvkysgmqlfvvksllleh 416

```

RESULT 13

AA069977
ID AA069977 standard; Protein: 861 AA.

```

XX AC AA069977;
XX DT 19-NOV-1998 (first entry)
XX DE Staphylococcus aureus histidine kinase.
XX KW Histidine kinase; KdpD; prevention; treatment; diagnosis; vaccine.
XX OS Staphylococcus aureus.
XX PN EP863208-A2.
XX PD 09-SEP-1998.
XX PF 17-FEB-1998; 98BP-0301167.
XX PR 25-FEB-1997; 97US-0039478.
XX PA (SMIK ) SMITHKLINE BEECHAM CORP.
XX PA (SMIK ) SMITHKLINE BEECHAM PLC.
XX PT Wallis NG.
XX PS Claim 14; Fig 2; 31pp; English.
XX CC The present sequence represents a Staphylococcus aureus histidine kinase.
XX CC The protein is related, by amino acid homology to the Escherichia coli
XX CC Kdp histidine kinase protein. The protein can be used to treat an
XX CC individual in need of histidine kinase, while the antagonist is used to
XX CC inhibit histidine kinase. DNA encoding and expressing histidine kinase
XX CC are used in immunological compositions. The active agents are useful
XX CC for preventing, treating, diagnosing and vaccinating against infections
XX CC of the upper and lower respiratory tract, cardiac disorders,
XX CC gastrointestinal disorders, CNS disorders, eye disorders, kidney and
XX CC urinary tract disorders, skin disorders, and bone and joint disorders.

```

CC They can also be used for assaying genetic variation and raising an
CC immunological response against e.g. Staphylococcus aureus.

XX Sequence 861 AA:

Query Match 6.6%; Score 143; DB 19; Length 861;
Best local similarity 21.2%; Pred. No. 0.00081;
Matches 109; Conservative 79; Mismatches 155; Indels 172; Gaps 25;

```

QY 17 TVLFLFLFESKSNVTLKSKELTFST--SNFLMIATVNVNLF-----YPAE-- 64
      |||||
Db 414 tlllmflfiig-----llslwtrsfllgflaainvtfvnyfitepryfevy 462
      |||||
QY 65 ----PYFIALSIYLNRONSLNLFYGLLPVASSDFERRAIRFIIDGTQGIWSSII 120
      |||||
Db 463 rfdypitfti-vslsslltsalqlkqtkqysltkkqjyrcdlilqfnd-----slk 512
      |||||
QY 121 TTYMIEFAGIALSYFLFSVFNVDIGRLKDSLTKMKYKKRLIPMNI--TMLVLYLIQVLYVI 180
      |||||
Db 513 qtyvevnlhmagyqf-----nqllgqslt-----lyvl 541
      |||||
QY 181 ESYNVIPLEKFRKV-----VIVYLI-----LFLILIS----- 208
      |||||
Db 542 ngskviktprlqnbndntqghegalvswlknereqagatctfpglnkwlipigtspikg 601
      |||||
QY 209 -FLSQYTKQKQVON-----EIMAKQEQIRNITQVSQOIESLYKD-----IRSF 250
      |||||
Db 602 llaigyssqvinpydaslllesmlnelavenvlllqtrlesnlqgerqlthsnflra 661
      |||||
QY 251 RHDYLNITLSRLGIE-----NKDLASIEKTYHQILEKTGHOLDTRY-----NIGHLA 299
      |||||
Db 662 shdirtplttlmgndilvshskdmstlek--eqllv---hsfigeslyllyvtnhlsl 716
      |||||
QY 300 NIQNDAYKGIISA-----KILEAQNKKAIVNVEVSSKIQLPPEMLDPTITLST 348
      |||||
Db 717 klqssnvqiklqpylvselveidmlerlhlkkrilvsssvnqfihidskllqalfn 776
      |||||
QY 349 LCDNATE--AFESLNPEIOLAEFKKNGSIVFIION-----STKEKQIDVSKIFKENVST 401
      |||||
Db 777 llenavkhsctklnslryasyeq---lefavideqpgalsleegq-----klfeffytc 828
      |||||
QY 402 KGSNR-----GIGLAKVNHILEHYPKTSIQ 426
      |||||
Db 829 -gsnkfyfdnqkesmgllqlylvqtll-hkhsnldq 861

```

RESULT 14

AA037397
ID AA037397 standard; Protein: 909 AA.

```

XX AC AA037397;
XX DT 14-FEB-2002 (first entry)
XX DE Staphylococcus aureus cellular proliferation protein #1567.
XX KW Antisense; prokaryotic cellular proliferation protein;
XX KW antibiotic; antibacterial; drug design.
XX OS Staphylococcus aureus.
XX PN WO200170955-A2.
XX PD 27-SEP-2001.
XX PF 21-MAR-2001; 2001WO-US09180.
XX PR 21-MAR-2000; 2000US-191078P.
XX PR 23-MAY-2000; 2000US-206848P.
XX PR 26-MAY-2000; 2000US-207727P.
XX PR 23-OCT-2000; 2000US-242578P.
XX PR 27-NOV-2000; 2000US-253625P.

```


Best Local Similarity 19.7%; Pred. No. 0.00027;
Matches 84; Conservative 83; Mismatches 125; Indels 135; Gaps 20;

```
QY 36 KELLTFSISNFMIM-ATMNVNLFYPAEPYFIALSTYLNKRONSLSLNIFGLPVAS 94
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 5 kdtstlaessilyllfplaglfinevgpkwlyllsvivf-----sis---yillyvnh 55
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 95 SDFRRAITFFIIDGTGIVMGSSITTYMIEFAGIALSYFLSVFNVDIGRLKDSLTKM 154
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 56 nrl--ntlmfyll-----llyflicyfvsvhpmislfyafavpf-tfknyvkt 107
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 155 KAKKRLIPNNTMLLYLLIQVLYIESYVNIPTLKRKRVYIYLLIFLLISFLSQYT 214
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 108 atnlfillmliclillyllynnyfva-----mmvyyvvaslilmldnfkmk 153
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 215 KQVONEIMAQKEAOIRNITQYSSQIESLYKDIRSFRHDYINILTSRLGIENKDLASIE 274
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 154 nreygkel-aeknrhncf-----171
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 275 KIYHOLEKTGHQLODTRYNIGHLANIONDAVKGILSAKILEAONKTI-----AVNV--- 326
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 172 -laegerhrigqldnt--lghv--faslskseykylidadevckaelalainklsr 225
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 327 -----EVSSTKIOPE-WELDFITILSIICDNALFEAFESLNPEIQOLAFPKKNGSIVF 378
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 226 esinkvrelldavklpsfleeid--sirkvdkadidftfe--nkelagvlsplkgsmly 281
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 379 IIONSTKEKQIDVSKIFKENVSTKGSNNGIGLAKVNHILEHYPKT---SIQTSNHHHLF 434
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 282 ml---Cre-----ainnvikhanaskvnhgkLktvnnh---310
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 435 KOLLIRK 441
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 311 klllme 317
```

Search completed: July 30, 2002, 15:14:13
Job time: 1594 sec


```

Db 6 ILLVALYNGKIVIFEFKNGIGLTFDRIFKAFLLKFLGIIFFTPQFLAVSKYLSYFIE 65
QY 65 PLVFLALSTYLNKRONSLNIFPGLLPVASSDLFRRAIIFFLIDGT-QGIV--MGSSIT 121
   ||:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
Db 66 PLFGIGSLFLRLGPKKILFYGGLFPMILVELFYGVSFVLPFGIGIVDGNPFL 125
   ||:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
QY 122 TYMIEFAGIALSYFLSFVNDIGRLKDSLTKKMKVKRRLIPNITMLLYLLIQLVLYIE 181
   ||:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
Db 126 LIMIFVCFIVL--VELKMLDYDFTLRREPLDTGFSKSLTKINMAGAYLLVMQSLSYLE 183
   ||:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
QY 182 SYNVPLTKFRKRVYIYLLIFLLISFLSQYTKQKQVONEMAKQKQIRNTQYSCQIE 241
   ||:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
Db 184 YEQGIQSTVTRHLLVFLFELFPMGIRKLDLYLKEKLOEELNQEOTLRDMERYSRHIE 243
   ||:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
QY 242 SYKDIFSRHDYINILTSRLGIENKDLASIEKIHQILEKTHOLODPTRYNIGHLANI 301
   ||:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
Db 244 ELYKEIRSRHDYINILTSRLGIENKDLASIEKIHQILEKTHOLODPTRYNIGHLANI 303
   ||:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
QY 302 QNDVAKGILSAKILAEONKKIYAVNEVSSKIQLEPEMLDFTITLSDNMLEAFESL 361
   ||:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
Db 304 RDRALKSLAGFKIAREKNIVENVEPEIQEGMSLDFLTIVSLCDNAIEVSAES 363
   ||:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
QY 362 NPEIQAFKKKGSIVFIIONSTKQKQIDVSKIFKENYSTKGSNRGIGLAKVNHLEHP 421
   ||:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
Db 364 QPHVSIATLKNQAGETFIENSIREGIDISELFSFGASSKGEGRVGLTYMKIVESH 423
   ||:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
QY 422 KTSLOTSNHHHLEFKOLLII 440
   ||:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
Db 424 NINLNTTCQNOVROVLTV 442
   ||:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:

RESULT 2
US-08-879-941-2
: Sequence 2, Application US/08879941
: Patent No. 6268172
: GENERAL INFORMATION:
: APPLICANT: Wallis, Nicola
: TITLE OF INVENTION: NOVEL HISTIDINE KINASE
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dechert Price & Rhoads
: STREET: 997 Lenox Drive, Building 3, Suite 210
: CITY: Lawrenceville
: STATE: NJ
: COUNTRY: USA
: ZIP: 08543
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/879, 941
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Bloom, Allen
: REGISTRATION NUMBER: 29,135
: REFERENCE/DOCKET NUMBER: GM10021
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 609-520-3214
: TELEFAX: 609-520-3259
: TELEX:
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 446 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
US-08-879-941-2

```

```

Query Match 34.6%; Score 752; DB 4; Length 446;
Best Local Similarity 39.2%; Pred. No. 8,9e-58;
Matches 172; Conservative 86; Mismatches 169; Indels 12; Gaps 5;

QY 12 LTYLVTLV--FLPLFESKSVNTLSKKELTFSISNPLIMIAVT-----MVNVNLFYPAE 64
   ||:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
Db 6 ILLVALYNGKIVIFEFKNGIGLTFDRIFKAFLLKFLGIIFFTPQFLAVSKYLSYFIE 65
   ||:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
QY 65 PLVFLALSTYLNKRONSLNIFGGLPVASSDLFRRAIIFFLIDGT-QGIV--MGSSIT 121
   ||:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
Db 66 PLFGIGSLFLRLGPKKILFYGGLFPMILVELFYGVSFVLPFGIGIVDGNPFL 125
   ||:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
QY 122 TYMIEFAGIALSYFLSFVNDIGRLKDSLTKKMKVKRRLIPNITMLLYLLIQLVLYIE 181
   ||:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
Db 126 LIMIFVCFIVL--VELKMLDYDFTLRREPLDTGFSKSLTKINMAGAYLLVMQSLSYLE 183
   ||:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
QY 182 SYNVPLTKFRKRVYIYLLIFLLISFLSQYTKQKQVONEMAKQKQIRNTQYSCQIE 241
   ||:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
Db 184 YEQGIQSTVTRHLLVFLFELFPMGIRKLDLYLKEKLOEELNQEOTLRDMERYSRHIE 243
   ||:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
QY 242 SYKDIFSRHDYINILTSRLGIENKDLASIEKIHQILEKTHOLODPTRYNIGHLANI 301
   ||:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
Db 244 ELYKEIRSRHDYINILTSRLGIENKDLASIEKIHQILEKTHOLODPTRYNIGHLANI 303
   ||:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
QY 302 QNDVAKGILSAKILAEONKKIYAVNEVSSKIQLEPEMLDFTITLSDNMLEAFESL 361
   ||:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
Db 304 RDRALKSLAGFKIAREKNIVENVEPEIQEGMSLDFLTIVSLCDNAIEVSAES 363
   ||:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
QY 362 NPEIQAFKKKGSIVFIIONSTKQKQIDVSKIFKENYSTKGSNRGIGLAKVNHLEHP 421
   ||:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
Db 364 QPHVSIATLKNQAGETFIENSIREGIDISELFSFGASSKGEGRVGLTYMKIVESH 423
   ||:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
QY 422 KTSLOTSNHHHLEFKOLLII 440
   ||:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
Db 424 NINLNTTCQNOVROVLTV 442
   ||:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:

RESULT 3
US-09-747-116-2
: Sequence 2, Application US/09747116
: Patent No. 6348340
: GENERAL INFORMATION:
: APPLICANT: Wallis, Nicola
: TITLE OF INVENTION: NOVEL HISTIDINE KINASE
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dechert Price & Rhoads
: STREET: 997 Lenox Drive, Building 3, Suite 210
: CITY: Lawrenceville
: STATE: NJ
: COUNTRY: USA
: ZIP: 08543
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/747,116
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/879,941
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Bloom, Allen
: REGISTRATION NUMBER: 29,135
: REFERENCE/DOCKET NUMBER: GM10021
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 609-520-3214
: TELEFAX: 609-520-3259

```

TELEX:
INFORMATION FOR SEQ ID NO: 2;
SEQUENCE CHARACTERISTICS:
LENGTH: 446 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-747-116-2

Query Match 34.6%; Score 752; DB 4; Length 446;
Best Local Similarity 39.2%; Pred. No. 8,9e-58;
Matches 172; Conservative 86; Mismatches 169; Indels 12; Gaps 5;

12 LTPVYL--FLFLFSKSNVTLSSKELTSLISNPLIMAVT-----MNVNLFYPAE 64
16 ILXALVINKLIVIPFKNGIGLTFDRIFKFLKFLGLGIFTFQPLANSKILSYFIE 65
65 PLVFALISYLNKNSLSNIEFGLPVASDLFRRAIFFLDGT--OGIV--MGSSITT 121
66 PLFGIGLSFLLRGLPKLILIEFGLPMLIVLEFYGVSFVLPFGGIVDGDNPFL 125
122 TYMEFAGIALSTLFSVENVDSIGRLKSLTFMKVKKRLIPNITMLLYLLIOVLVIE 181
126 LIMFVCEYL--VELKMLDYDTRLRREPLDTGFKSLTKINMAGAYILVMOSTYLE 183
182 SYNVIPFLKPKKPVIVYLLFLILISLSQYTKOKVONEIMAOEAOIRNITQSOIE 241
184 YEGIOSTYVRHLIVLFYLLFFMGIGIKLDYTLKEKLELNOEOTLRDMERSRIIE 243
242 SLKYDISFPHDYLTITSLRIGENKDLASIEKIYHOILEKTHGLODTRNIGHLANI 301
244 ELKEIHSFPHDTNLTSLRGLIEEDMEQKEIYDVLNRSKIDNDKIDGLRVNI 303
302 QNDVAKGILSAKLEAKNKIAVNEVSSKIOLPEMELDTITLISLIDNAIEAEPESL 361
304 RDRALKSLAGKFIKAREKNIVNVEPPEIYEGSKLDFLIVSLIDNAIEVSAEAS 363
362 NPEIOLAFKFNKSIYFIIONSTKKEOIDVSKIFKENTSKGNSNGIGIAKVNHLEHP 421
364 QPHYSIAFLKNGAOEFTEIENSISKEGIDISLSEFSGASSKKEBVGGLYTYMKIVESH 423
422 KTSLSQTSNHHHLFKOLLIT 440
424 NFINNTTCQNOVFOYLV 442

RESULT 4

US-08-858-207A-327

Sequence 327, Application US/08858207A

Patent No. 6348328

GENERAL INFORMATION:

APPLICANT: Black, Michael

APPLICANT: Hodgson, John

APPLICANT: Knowles, David

APPLICANT: Stodola, Robert

TITLE OF INVENTION: No. 6348328e1 Compounds

NUMBER OF SEQUENCES: 552

CORRESPONDENCE ADDRESS:

ADDRESSEE: SmithKline Beecham Corporation

STREET: 709 Swedeland Road

CITY: King of Prussia

STATE: PA

COUNTRY: USA

ZIP: 19406-0939

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS for Windows Version 2.0

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/858,207A

FILING DATE: 09-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/017670
FILING DATE: 14-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimm, Edward R
REGISTRATION NUMBER: 38,891
TELEPHONE/DOCKET NUMBER: P50475
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 327;
SEQUENCE CHARACTERISTICS:
LENGTH: 108 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 6348328e
US-08-858-207A-327

Query Match 11.9%; Score 258; DB 4; Length 108;
Best Local Similarity 48.1%; Pred. No. 1.6e-15;
Matches 50; Conservative 22; Mismatches 32; Indels 0; Gaps 0;

337 MELDFITLSTLIDNAIEAFAESINPEIOLAFKFNKSIYFIIONSTKKEOIDVSKIRK 396
1 MSLDFITVSLIDNAIKASASOPHVSIAFLKNGAOEFTEIENSISKEGIDISEIRS 60
397 ENSTGKSNRGIGLAKVNHLEHPKTSLSQTSNHHHLFKOLLIT 440
61 FGASSKKEBVGGLYTYMKIVESHENFINNTTCQNOVFOYLV 104

RESULT 5

US-08-858-207A-326

Sequence 326, Application US/08858207A

Patent No. 6348328

GENERAL INFORMATION:

APPLICANT: Black, Michael

APPLICANT: Hodgson, John

APPLICANT: Knowles, David

APPLICANT: Stodola, Robert

TITLE OF INVENTION: No. 6348328e1 Compounds

NUMBER OF SEQUENCES: 552

CORRESPONDENCE ADDRESS:

ADDRESSEE: SmithKline Beecham Corporation

STREET: 709 Swedeland Road

CITY: King of Prussia

STATE: PA

COUNTRY: USA

ZIP: 19406-0939

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/858,207A

FILING DATE: 09-MAY-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/017670

FILING DATE: 14-MAY-1996

ATTORNEY/AGENT INFORMATION:

NAME: Gimm, Edward R

REGISTRATION NUMBER: 38,891

REFERENCE/DOCKET NUMBER: P50475

TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-4478

TELEX:
INFORMATION FOR SEO ID NO: 2:

TELEX:
; INFORMATION FOR SEO ID NO: 2:
;

```

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 618

```

TYPE: PRF
ORGANISM: Staphylococcus warneri
US-09-299-378-4

Query Match 5.4%; Score 117; DB 4; Length 618;
Best Local Similarity 21.4%; Pred. No. 0.038;
Matches 116; Conservative 76; Mismatches 187; Indels 164; Gaps 27;

```

QY 8 LNSGLTLYTVL-----FLLEFSKSVNVTLSKKELELFSSNLI-----MAY 52
DB 92 LNTLTAKKVLDPACGCGISIAFIEIFNRKISKSIVDF-INNVAVIIVDSNEKINF 150
QY 53 TAWN-----VNLTPAEP-LYFI-----ALSTYINRONSLSN-----84
DB 151 TKINILTMVYLYNDISIFDKVKPNIAIDFVQEKKEFELFMFNSQOMNFINDNFEN 210
QY 85 --IFGGLPVASSDLF--RRAT-----IFELIGTQOG 112
DB 211 FDIYIGNPPYVS--LYGRRAINKSEDKRQFLIRNYDFIPKNNKNGKENTYMFELNGLQL 268
QY 113 IVMGSSI-----ITTYMIEFAGIALSYLFLSVFN-----VDIGRLKDSL 151
DB 269 LKNGTLFIFYDITLLESSFESIRKYLETAIIKOLDINKAFSDVYSCQIISLTKNAS 328
QY 152 TKMKV-----KKRLIMNITMLL---YLLI---QVLVIES--YNAVPTLK--PRKF 194
DB 329 NKEAIVSIKDMONNTIOINODIMLHDKFYEPNISDKKINSILEKYNNKSDDELQYTFPPK 388
QY 195 VVIYVLLIFLISFLSOYTRQKOVNELMAQKEAQRNITOYSOIES-----LYK 245
DB 389 ELRTSTMLNMESSEFVKDY-KPEIDFHVMPYYKC-AKNLSEPFQNMHSNHYFYDALOK 446
QY 246 DIRSRHYL---NLTLRLIGIENKDLASIEKIY-----HOLEKTHGLODTRNIGH 297
DB 447 KINDSLHEELKKKIKKKRIGLGNLEFKKPKLFIRQSANKLATDGMASNSNLIY 506
QY 298 LANIONDAVKGILSAKILEAQNKKIAVNEVSSKIQLPEMELLDITILSLCDNAIEAA 357
DB 507 LSKATND-IKIDIMLKIKTCAOLNS-----ELLTFIAL-----TNRIIRK 544
QY 358 FESLMEIOLAFKKNGSIVFIIONSTKEKOIDVSKIFKENEYSGKGNRIGIAGVNHIL 417
DB 545 ASGKQOLKIDLK---TIPLCFNEINSKLL---IFASNATKKNNELSSLEKINQII 597
QY 418 EHY 420
DB 598 YKY 600

```

RESULT 10
US-08-963-901-2
Sequence 2, Application US/08963901
GENERAL INFORMATION:
APPLICANT: Wallis, Nicholas G.
APPLICANT: Traini, Christopher M.
APPLICANT: Kosmatka, Anna L.
APPLICANT: Shilling, Lisa K.
APPLICANT: Warren, Richard L.
TITLE OF INVENTION: NOVEL HISTIDINE KINASE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESS: Dechert Price & Rhoads
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/963,901
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Dickinson, Todd O
REGISTRATION NUMBER: 28,354
REFERENCE/DOCKET NUMBER: GM10058
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-994-2252
TELEFAX: 215-994-2222
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 457 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-963-901-2

Query Match 5.2%; Score 114; DB 4; Length 457;
Best Local Similarity 19.6%; Pred. No. 0.045;
Matches 96; Conservative 76; Mismatches 167; Indels 152; Gaps 24;

```

QY 38 LTFESISNELLMAVTAVNVLLEY-----PAEPLFYALSI 73
DB 9 IAVISITVELSALISFVLNVVYHMLKASNDAKIMTKLEARQESAKPTHIQOYFK 68
QY 74 YLNRONSLSNI-----FYGLLPVASSDLFPRAT-----F 104
DB 69 HLGQNVYIMVVDQKGRKTFG--EPREDLSQNALNNVNMNDYGINDKPFALEVTF 127
QY 105 F-IIDGTQIGVWGSSIIITYMIEFAGIALSYLFLSVFNVDIGR-LKDSLTKMKYKRLI 161
DB 128 FDNVTDNVGI-----NEKTKDGSIAV-----FMRDIOETFSERTFLAV-----168
QY 162 PMNITMLLYLLIOLVLYIESVNIPTLKFRKFVYIYLLIFLISFLSOYTRQKOVNE 221
DB 169 ---LMLLFLSIS-LVIASYSIIRVKKIK-----LATELLIDGDETPRKQTRKE 218
QY 222 I-----MAQKEAQRNITOYSQO-----TESLYKDI-----RSFRHLYL 255
DB 219 IGTLOYHFNKMRRESLQYDQROHFVONVSHETKPTLTHIHLLSELQOTSOKTLRQOYI 278
QY 256 N---ILTLRLIGIENKDLASIEKIYHQLKKTGHLODTRNIGHLANIONDAVKGILSA 312
DB 279 NDIYITITQLSGLTTELLLSLSDNHQ-----HLFPDKIOVDL-----IKDIINH 325
QY 313 KILEAQNKKIAVNEVSSKIQLPEMELLDIT-----TILSLCDNAIEAAEFESLNPEI 365
DB 326 EQFADEKSLIY-----LADLESINFLGNORLLHQAISMLINAIK--YTDVGAI 374
QY 366 QLAFFKNGSIVFIION--STKEKOIDVSKIFKENE--STKSGNRIGIAGVNHILE-HY 420
DB 375 DALDHSNNIIFTISNDGSPISPOAE-ARLEFEKYVSHDNSNGUGALITKSIETLHH 433
QY 421 PRTSLQTSNHH 431
DB 434 GTIOFTQSNEX 444

```

RESULT 11
US-08-568-985-2
Sequence 2, Application US/08588985
Patent No. 5777094
GENERAL INFORMATION:
APPLICANT: Michiyuki MATSUDA et al.
TITLE OF INVENTION: CDNA OF DOCK180 GENE AND DOCK180 PROTEIN


```

: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Wenderoth, Lind & Ponack
: STREET: 805 Fifteenth Street, N.W., #700
: CITY: Washington
: STATE: D.C.
: COUNTRY: U.S.A.
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: Wordperfect 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/588,985
: FILING DATE: January 19, 1996
: CLASSIFICATION: 356
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Warren M. Cheek, Jr.
: REGISTRATION NUMBER: 33,367
: REFERENCE/DOCKET NUMBER:
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-371-8850
: TELEFAX:
: TELEX:
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1865 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEICAL:
: ANTI-SENSE:
: FRAGMENT TYPE:
: ORIGINAL SOURCE: spleen cell of homo sapiens
: US-08-588-985-2

```

```

Query Match          5.2%; Score 113; DB 1; Length 1865;
Best Local Similarity 20.1%; Pred. No. 0.39;
Matches 95; Conservative 90; Mismatches 208; Indels 80; Gaps 21;

QY 21 LLEFESKVSNTLSKKELEFISNFIIMIAVTMNVNLFYPAEPIYFALSITYN-RQ 78
DB 772 LLOLFRSINDMSSMSDQTV-RVKGAAKLYPTIVNDVKLVDPKELSKMTEFLINVP 830
QY 79 NSLSLNIFFGLPVAASDLFRRAIFFLIDGTGQIVMGSSITTYM-----IFRAGIALS 133
DB 831 GLLTIOKLYCLIEIYHSDLFTHDCREIL-----LPMMTDQKYLHERODELACCOLLS 885
QY 134 YLEFSEVNDIGRLKSLTKMKVKKRLIPMNITML-----LYLLIQVLYVIES 182
DB 886 HILEVLYRKVDGPTQRHV-QIIMEKLRVTNRTVISMGDSLEIGNFVACMTAILRQMED 944
QY 183 YNVIPTLK-FRFEVVIYVLIIFLILISFSQTKO-----KVQNEIMAKREAQIRNITO 235
DB 945 YRAHALTKITGKMTDVVDFLMEFTIMERNLIGKNVYPPDWVIM--VWQNKVFRLAINQ 1002
QY 236 YSQOIESLYADIRSF---HDYLINI---LTSRLGIENKDLASIEKI---YHOILEKT 284
DB 1003 YADLANKKFLDQANFELQJLNNYFHLAVAFLOESLOJENFSSAKRAKILNKYGDMRROI 1062
QY 285 GHLODTRKYIG-HLANIQNDANKGLSLAKIL-EAONKRIANV-----EVSSTIQ 333
DB 1063 GFELRDMMYLGQHKIKFIPENAGPLLEMTLIPETELKRATPIFFDMACCEHSTNSEQ 1122
QY 334 LPEMELDFTTILSLCDNAIEAA---FESLNPEIQO-----AFKKGSIYFIION 382
DB 1123 MFEENEI---ITKL---DHEVEGGRGDEQYKVLFRKILHECRKHKYLAKGETFVKLVV 1175

```

```

QY 383 STREKQDVSKIFKENVSTGNSNGIGLAKVNHLEHYPTSLQTSNHHLEK 435
DB 1176 RIMEFLIDYRTIMHD-----ENKENMSCTVNLNFKYEIEREMYYIRLYLK 1222

RESULT 12
US-08-971-988-2
: Sequence 2, Application US/08971988
: Patent No. 5786461
: GENERAL INFORMATION:
: APPLICANT: MICHIOKI MATSUDA et al.
: TITLE OF INVENTION: CDNA OF DOCK180 GENE AND DOCK180 PROTEIN
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Wenderoth, Lind & Ponack
: STREET: 805 Fifteenth Street, N.W., #700
: CITY: Washington
: STATE: D.C.
: COUNTRY: U.S.A.
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: Wordperfect 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/971,988
: FILING DATE: 17-NOV-1997
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/588,985
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Warren M. Cheek, Jr.
: REGISTRATION NUMBER: 33,367
: REFERENCE/DOCKET NUMBER:
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-371-8850
: TELEFAX:
: TELEX:
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1865 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEICAL:
: ANTI-SENSE:
: FRAGMENT TYPE:
: ORIGINAL SOURCE: spleen cell of homo sapiens
: US-08-971-988-2

```

```

Query Match          5.2%; Score 113; DB 1; Length 1865;
Best Local Similarity 20.1%; Pred. No. 0.39;
Matches 95; Conservative 90; Mismatches 208; Indels 80; Gaps 21;

QY 21 LLEFESKVSNTLSKKELEFISNFIIMIAVTMNVNLFYPAEPIYFALSITYN-RQ 78
DB 772 LLOLFRSINDMSSMSDQTV-RVKGAAKLYPTIVNDVKLVDPKELSKMTEFLINVP 830
QY 79 NSLSLNIFFGLPVAASDLFRRAIFFLIDGTGQIVMGSSITTYM-----IFRAGIALS 133
DB 831 GLLTIOKLYCLIEIYHSDLFTHDCREIL-----LPMMTDQKYLHERODELACCOLLS 885
QY 134 YLEFSEVNDIGRLKSLTKMKVKKRLIPMNITML-----LYLLIQVLYVIES 182
DB 886 HILEVLYRKVDGPTQRHV-QIIMEKLRVTNRTVISMGDSLEIGNFVACMTAILRQMED 944
QY 183 YNVIPTLK-FRFEVVIYVLIIFLILISFSQTKO-----KVQNEIMAKREAQIRNITO 235

```



```

Db      193 GIELLHF-SVIDKDEREYRCSDEEGSEDSYQPLFONDPPAKSIVVHPFGKD- 250
      168 LLYVLLIOVLVIESYN-VIPLFKRFRVYVLLIFLLISFISOYTKOKV--NEIMA 224
      251 -----YIPDSVSPFIPSMIF-TFVLLITFIFTIYIFRQKLTLEKNDPNNMTH 299
      225 OKKAOIRNTQYSGQIE--SLYKDIRSPRHVDNLNLTSLRGIEKDLASIEKTYHOLE 282
      300 EFKTPISTISLAQMLKDPVAKSPQMFQH-----ISGVINDETKRLRFQVEKVLQ 350
      283 KTGHOLODRYRNIGHLANTQNDVAKGLS--AKILEAONKRIAVNEVSSK-IQLEPEMEL 339
      351 MSMEFROKATLKKKELD--ANELISGVINTFALKVERYNKITSNLEATNPFVFADEMI 408
      340 LDFITLITLQDNATIAEESINPEIQALF--FKKNGSIVFIQ--NSTEKQIDVSKI 394
      409 TN--VIFNMDNAVKY--KPEDLVLDVYTWNNEPKLMTISIQDNGIGIKKENLKV 461
      395 FKENYSTKGSNR-----GIGLAKVNHILEHYPKT 423
      462 FDKFYRVHTGNLHDVYKFGSLAYVKIITQDKGT 496

RESULT 15
US-09-081-686-2
Sequence 2, Application US/09081686
Patent No. 6162619
GENERAL INFORMATION:
APPLICANT: Wallis, Nicola G.
APPLICANT: Shilling, Lisa K.
APPLICANT: Wang, Min
APPLICANT: Jaworski, Deborah D.
APPLICANT: Ingraham, Karen A.
APPLICANT: Yigong Ge, James J.
APPLICANT: Holmes, David J.
APPLICANT: Zalacain, Magdalena
APPLICANT: Throup, John
APPLICANT: Biswas, Sanjoy
TITLE OF INVENTION: Histidine Kinase
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Decnet, Price & Rhoads
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2793
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/081,686
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/048,339
FILING DATE: 30-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: Falk, Stephen T
REGISTRATION NUMBER: 36,795
REFERENCE/DOCKET NUMBER: GM10006
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-994-2488
TELEFAX: 215-994-2222
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 446 amino acids
TYPE: amino acid
STRANDEDNESS: single

```

```

; TOPOLOGY: linear
US-09-081-686-2
Query Match          5.0%; Score 108; DB 4; Length 446;
Best Local Similarity 18.3%; Pred. No. 0.15;
Matches 96; Conservative 78; Mismatches 133; Indels 218; Gaps 23;

      36 KELTFSISNFIIMAVMVNANLFPAEPIYFALSTIYNRONSLSNIFGLVASS 95
      6 KLLTSPFNRSEFRIIGVGLVTHAIYLFPEYIQLF--GEKFNSARVETELKTS 62
      96 DLFRAIIFIIDTQGIWMSIITTYMIEFAGIALSY--LFLS----- 138
      63 D-----EIPSLQSYKSLISAKLDIVDKRLPL 93
      139 VFNVDI--GRKDSLTAKKVKRKLIPNFI--TMLLYILLIOVLVIESYNTPLKRFY 195
      94 VHDLDIKDGLSNYIV--MIDMSVSTADGQVTVQVGVGVVY--KKAANI 140
      196 VIVYL-ILFLILISF--LSQYTKOKVONEI-----MAQKQAIR-----NITQ 235
      141 LLLYPTFELVYIASFVSYFYTKRLNPLFYISEVTSKQDLDDNIRFDESRKDEGE 200
      236 YSQIIESLYKD-----IRSRHDYLNLTSLRGIE-- 267
      201 VGRQINGMYEHLKYIHELESRENOIVKIQKVSFVGASHETKPLASLRILLENMCH 260
      268 -----KDLASIEKTYH--QILEKTGHQ-----LQD 290
      261 NGDYKDHPRKYIAKSINKIDQMSHLEVELESKFQEWTECRETLVYKPYVDILSRGE 320
      291 TTYNIG-HLANTQNDVAKGLISAKILEONKRIAVNEVSKIDLPMELEDDITLSTL 349
      321 LAHSGVTIENQLDATVAVMSLRALD-----KVLTNL 353
      350 CQNAIEAAEESINPEIQALFAFKKNGSIVFIQ--STKE-----KQIDVSKIFKENY-- 399
      354 ISNAIK-----YSDKNGRVYIISQDGYLSIKNTCAPLSQDEHLEHFDLFYHS 400
      400 ---STKGSNRGIGLAKVNHILEHY-----KTSLOF 427
      401 QIVYDKDESSGGLYIVSNLLESTOMDYRYLPEHGMFKISLOT 445

```

Search completed: July 30, 2002, 15:14:44
Job time: 520 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:15:18 ; Search time 29.21 Seconds
(without alignments)
1450.715 Million cell updates/sec

Title: US-09-833-017-4
Perfect score: 2173
Sequence: 1 MNEALMILSNGLTLYLVLE.....KTSIQTSNHHLPKOLLIIK 441

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR_71:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query length	DB ID	Description
1	755	34.7	446	2	H97929 histidine kinase (
2	749	34.5	446	2	D95061 sensor histidine k
3	455	20.9	447	2	F96909 histidine kinase-1
4	427.5	19.7	441	2	B95261 probable sensor hi
5	423.5	19.5	441	2	G98126 histidine protein
6	418.5	19.3	431	2	AC1081 sensor histidine k
7	416	19.1	431	2	AD1438 sensor histidine k
8	379	17.4	423	2	S58480 aqrc protein - Sta
9	372	17.1	423	2	D89995 accessory gene reg
10	362	16.7	428	2	G56273 protein-histidine
11	348	16.0	448	2	S57909 probable histidine
12	223.5	10.3	452	2	B97095 membrane associate
13	184.5	8.5	433	2	AB1533 two-component sens
14	174	8.0	542	1	E66600 two-component sens
15	171.5	7.9	532	2	E83699 two-component sens
16	163.5	7.5	657	2	D81308 Probable Integral
17	155.5	7.2	440	2	C55521 vits protein - Clo
18	154	7.1	533	1	A70009 two-component sens
19	151.5	7.0	921	1	G71705 alkaline phosphata
20	151	6.9	543	1	D65222 hypothetical 60.6
21	151	6.9	543	2	C91267 two-component sens
22	151	6.9	543	2	H86107 probable 2-compone
23	150.5	6.9	543	2	AH1023 two-component sens
24	148.5	6.8	473	2	G97085 sensory transducti
25	146	6.7	1447	2	F82909 hypothetical prote
26	145.5	6.7	442	2	C95070 sensor histidine k
27	145.5	6.7	442	2	A97938 vncs, histidine ki
28	143	6.6	538	2	G84129 two-component sens
29	143	6.6	885	2	D90000 sensor protein kdp

30	142	6.5	824	2	A86783 glycosyl transfera
31	138.5	6.4	363	2	B89907 hypothetical prote
32	137	6.3	428	2	S32935 kinb protein - Bac
33	136.5	6.3	1008	2	F71727 acriflavin resista
34	136	6.3	733	2	C89940 hypothetical prote
35	136	6.3	754	2	G70124 hypothetical prote
36	135.5	6.2	542	2	E90604 hypothetical prote
37	135.5	6.2	569	2	B70113 hypothetical prote
38	134.5	6.2	553	2	AD0579 sensor kinase dph
39	134	6.2	565	2	G64532 methyl-accepting c
40	133.5	6.1	538	2	D82180 probable sensor ki
41	131	6.0	564	2	F71975 methyl-accepting c
42	130.5	6.0	714	2	C90100 hypothetical prote
43	130	6.0	590	2	G90127 hypothetical prote
44	129.5	6.0	339	2	C96927 sensory transducti
45	129.5	6.0	532	2	H83993 two-component sens

ALIGNMENTS

RESULT 1
H97929 histidine kinase (EC 2.7.3.-) [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #ext_change 02-Nov-2001
C:Accession: H97929
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S. Y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A978/2; M01D:21429245; PMID:11544234
A:Accession: H97929
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-446 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK9268.1; PID:q15458033; GSPDB:GN00174
C:Genetics:
A:Gene: hki3
C:Keywords: phosphotransferase

Query Match	34.7%	Score 755;	DB 2;	Length 446;
Best Local Similarity	39.1%	Pred. No. 9.5e-39;		
Matches 175;	Conservative 88;	Mismatches 171;	Indels 14;	Gaps 5;
QY	1	MNEALMILSNGLTLYLVLEFLFESKVSNTSKKELTFSINPLIMAVT-----MV	55	
DB	1	MAINMILKALVINGLETIV---IFPKVDGIGLFDRIKFAFLKFLGLITFTTPOFLAV	56	
QY	56	VNVLTPAEPLVIALSLYLRNONSLSINIFYGLLPVASSDERRAIIFILDGT-OGIV	114	
DB	57	SKYLSFTEPLFEGISLFLRGLPKKILIFGJFPMILVELFEGVSYVLPFLQGIIV	116	
QY	115	-MGSIIITTVIERFAGALSYFLSVPNVDIGLTKSLKMKVKKRLPMMNTMLLYL	172	
DB	117	DGDGNPILFLIMFVCFYVL--VELKMDYDPTFLRKREPLDTGFSQSLTIINMAAYIL	174	
QY	173	LIOVLVIESVNPITLFRFVVIVYLIFLLISFLSYTRKQVONEIMAKQEOINR	232	
DB	175	VMSQSYLEVEEGIOSTVRLHILVFYLLFPMGSIKKLDLYLEKLOEELNOCOTLRVD	234	
QY	233	TOYSOQIESLYKDIRSRHDYLNITSLRGINKDLASTIKRYQILEKTHOLODR	292	
DB	235	MERYSHHELYKETRSRHDYTNLISLRIGIEEDMEQIKELYSDVLRDSOKLODK	294	
QY	293	VNIGHLANTQNDVAKGISANILEKONKKAIVNVESSKIQLPBMLDFTTILSLCN	352	
DB	295	YDGLRVNTRDALSKSLAGFKIKAREKNIVNVEPEIOVEMSLDPLTIVSILCN	354	
QY	353	ALFAAFESLNPEIQAFPKKNSIYFIIONSTKQKQIDVSKIFRENVSTGSRNGIGLAK	412	

Db 355 ALEASAEAPHVSAFLAKAGQETITENSKEEGIDISELFGASCASKGEEGVLGT 414
QY 413 VNHLEHYPKTSLOTNSHHHFKOLLIT 440
Db 415 VMKIVESHPTNLTTCQNOVFQVLT 442

RESULT 2
D95061
sensor histidine kinase B1pH, probable (imported) - Streptococcus pneumoniae (strain TIGR
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: D95061
R:Rettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: D95061
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-446 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK74685.1; PID:g14972002; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP0527

Query Match 34.5%: Score 749; DB 2; Length 446;
Best Local Similarity 37.8%; Pred. No. 2, 2e-38;
Matches 170; Conservative 89; Mismatches 173; Indels 18; Gaps 4;

QY 1 MNEALMIL-----SNGLLTYLVLEFLFESKSVNTLS-----KKELLFESISNFLMIA 51
Db 1 NMIAWILLTYLVNGL-----EIVIFKVDGIDLTFRKFAKLTLAFAVYVMS 52
QY 52 VMVAVNVFPAEPLFYFALSIVLRONSLSLNFYGLPLVASSDFERRAIFELIDGT- 110
Db 53 YIVGVNVLSYEMEPYIGLSFLLRGLPKKLFYGLFPMILVNFYGVGYFVPLFG 112
QY 111 OGIVNGSSITTYMTERAGALSTFTSVFNVDIGRLKDSLTKMKYKKRLPMNTMLY 170
Db 113 OGQVVDGYSTFGCLITIFNFELSLAFKMDYDFTSLRKLIDKAPQKSLTQIMMGY 172
QY 171 YLLIOVLVIESYNYIPTLKRFVIVLYLFLILISFLSYTKOKOVONEIMAKQEI 230
Db 173 YIVMSLSFFFEYEGSIOKTVRHILVFLYLLFPMGVAKKIDYLRKEXLERLEQALRY 232
QY 231 RNITQYSGQIESLYKDIRSFHDYLNLTSLRGLGIEKDLASIEKTYHQILEKTHOLD 290
Db 233 RMERYSRHIEELKVEVSFRHDYNTLTSLRGLGIEEDMDQKVEYGVGLKDDSKOLN 292
QY 291 TYRNLGHLANIONDAVKILSAKILEAQNKKIAVNEVSSKIQLPMELDLFTILSLIC 350
Db 293 NKYDGLRVNIRDAKAKSLAKGLKARKNIVNVEPELQVEGMSLDELITVSLIC 352
QY 351 DNAIEAFAESLPEIQLAFAFKKNGSYVFTIONSTKEQIDVSKIFENSTGSGNIGL 410
Db 353 DNAIEAFAESQPHVSAIFLNGAQEFIIENSKEGIDISEIFSGAGSGEERGVL 412
QY 411 AKVNHLEHYPKTSLOTNSHHHFKOLLIT 440
Db 413 YVMKIVESHPTNLTTCQNOVFQVLT 442

RESULT 3
P96909
histidine kinase-like Arpase [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001

C:Accession: P96909
R:Rolling, J.; Brenner, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: P96909
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-447 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK78065.1; PID:g15022902; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC0080

Query Match 20.9%: Score 455; DB 2; Length 447;
Best Local Similarity 27.4%; Pred. No. 1, 6e-20;
Matches 124; Conservative 96; Mismatches 163; Indels 70; Gaps 10;

QY 11 GLTYLVLEFLFESKSVNTLSKKELLFESISNFLMIATVMVNVNFPYAEPLFYFA 70
Db 33 GMIVTEIISLIFLM-----LGMQVTVMA-----YLI 61
QY 71 LSTYLRONS---LSLNFYGLPLVASS-----DLERRAIFELIDGTQIGVMSSTI- 120
Db 62 PCLFMWLTSDVLSIS-----LTVAGCLIIYWDYFLSVLYVFLVQSVYKNDVSIH 116
QY 121 -TYMTERAGALSTFTSVFNVDIGRLKDSLTKMKYKKRLPMNTMLYLLIOVLY 179
Db 117 WTFPIEFYIGYVYSRTTKKVRD-----KFKVYNGHLGTTGVVAAGTLIMLFV 167
QY 180 ISYVNY-----PLKRFKRVIV-YLILFLLISFLSYTKOKOVONEIMAKQEI 228
Db 168 FNTNVNVPAPNNSVYNSMTMRKGVLFESYAILLIIVITRLGILKEME--LKSKE 224
QY 229 QIRNITQYSGQIESLYKDIRSFHDYLNLTSLRGLGIEKDLASIEKTYHQILEKTHOL 288
Db 225 EFQSLQDYTNKLEKDKMRGFRHDYINLTLSMAGYTONDLEPFDOKIMPLSKAM 284
QY 289 QOTRWIGHLANIONDAVKILSAKILEAQNKKIAVNEVSSKIQLPMELDLFTILSI 348
Db 285 KSNMFKIGLQNIPELPEIKGMFSAKIIRAQGTGIDYIDVAESIKSFNHEIDLSVIGI 344
QY 349 LCDNAIEAFAESLPEIQLAFAFKKNGSYVFTIONSTKEQIDVSKIFENSTGSGNIGL 408
Db 345 LIDNAIEASEKCDRSMKVAIVNDRSVMIIVINNVEELIPYIKLYKGFSTKGDNRI 404
QY 409 GLAKVNHLEHYPKTSLOTNSHHHFKOLLIT 441
Db 405 GLSNLKDITGKRPVMDTVIEDNQFKIIDIK 437

RESULT 4
B95261
probable sensor histidine kinase Comp [imported] - Streptococcus pneumoniae (strain T
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: B95261
R:Rettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: B95261
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-441 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK76283.1; PID:g14973747; GSPDB:GN00164; TIGR:
A:Experimental source: strain TIGR4
C:Genetics:

Db 197 IYIVTVVINTAFNELKVN---OKE-QLBQLQDYVTLTSLHREKRVFRHDYVNLSTL 251
OY 262 RLGLENKPLASIEKTYHOILEKTHQLODTRVYNGHIANIQNDAVNGILSAILLEAONKK 321
Db 252 VGYINDNMPIKTYFEENNYIPINKTIESNNYKISLQNNHYTELKGLAVLIRAOELK 311
OY 322 IAVNVEVSSKIQLEPMELDFITLISLQNAIEAFESINPEIOLAEPFKNGSIYFIIQ 381
Db 312 IDALEVEVPEIDKISMSIDCKYVGLLDNAVEALTCENPVIKIAFVKGDSIIIVFA 371
OY 382 NSTEKQIDVSKIREKENTSTGSRGIGLAKVNHILEHPKTSLOQ 427
Db 372 NSLPVNMPIYKIFEEGSGTKGEGRGIGLASLREIMKKYSHVALDT 417

RESULT 7
AD1438
sensor histidine kinase (AgrC from *Staphylococcus*) homolog lin0043 [imported] - *Listeria*
C:Species: *Listeria innocua*
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AD1438
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloechet
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entlan, K.D.; Fsihl, H.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
Ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A.; Title: Comparative genomics of *Listeria* species
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AD1438
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-431 <GRL>
A:Cross-references: GB:AL597022; PIDN:CA095276.1; PID:g16412464; GSPDB:GN00178
A:Experimental source: strain Clp11262
C:Genetics:
A:Gene: lin0043

Query Match 19.18; Score 416; DB 2; Length 431;
Best Local Similarity 26.28; Pred. No. 3.7e-18;
Matches 126; Conservative 94; Mismatches 148; Indels 112; Gaps 16;
OY 21 LILFESKYSNNTLSKKEIT--LESISNPLIMAVTMVNNVLEFPAP--PLYFIAL-----S 72
Db 4 IIMATIQITGIFIAIOLITKRVFSIKGVLVIAIAMAFLPLVGYVMSMIFVLIVFVSA 63
OY 73 IYLNQON-----SLSNIFYGLPVAASDLFRRAIIFLIDGTQIGVMSIITTYMIEF 127
Db 64 LWMKKNVIVSGSITLVYII-LTISDS-----IVGFIL-----VPG----- 99
OY 128 AGIALSYFLSVFVNDIGRLKDSITKMKVKRLIP-----NITMLLYLLIOVLYI 180
Db 100 ---INFKYDELFN-----ELPLTLYCAGMLANLIVFSFLKRL-I 136
OY 181 ESYNVIPLTKFERKVVYVILLFILISF-----LSQYT--- 214
Db 137 EKVNISRFERHKKAYIIFISIVALTVAFYNNIYAGSIAGDSVLIKNTLLFTGYTILL 196
OY 215 -----KQVONEIMAKQEQIRITQYSQISLTKDRSRPHNYNLTSL 261
Db 197 IYIVTVVINTAFNELKVN---OKE-QLBQLQDYVTLTSLHREKRVFRHDYVNLSTL 251
OY 262 RLGLENKPLASIEKTYHOILEKTHQLODTRVYNGHIANIQNDAVNGILSAILLEAONKK 321
Db 252 VGYINDNMPIKTYFEENNYIPINKTIESNNYKISLQNNHYTELKGLAVLIRAOELK 311
OY 322 IAVNVEVSSKIQLEPMELDFITLISLQNAIEAFESINPEIOLAEPFKNGSIYFIIQ 381
Db 312 IDALEVEVPEIDKISMSIDCKYVGLLDNAVEALTCENPVIKIAFVKGDSIIIVFA 371
OY 382 NSTEKQIDVSKIREKENTSTGSRGIGLAKVNHILEHPKTSLOQ 427
Db 372 NSLPVNMPIYKIFEEGSGTKGEGRGIGLASLREIMKKYSHVALDT 417

Db 372 NSLPVNMPIYKIFEEGSGTKGEGRGIGLASLREIMKKYSHVALDTKVTNREVIOELIEM 431
RESULT 8
558480
agc protein - *Staphylococcus aureus*
C:Species: *Staphylococcus aureus*
C:Date: 06-Dec-1996 #sequence_revision 06-Dec-1996 #text_change 15-Oct-1999
C:Accession: 558480; S20797
R:Novick, R.P.; Projan, S.J.; Korblum, J.; Ross, H.F.; Ji, G.; Kreiswirth, B.; Vande
Mol, Gen. Genet. 248, 446-458, 1995
A:Title: The agc 82 operon: an autocatalytic sensory transduction system in *Staphyloc*
A:Reference number: 558478; MUID:96004766
A:Accession: 558480
A:Molecule type: DNA
A:Residues: 1-423 <NOV>
A:Cross-references: EMBL:X52543
A:Note: the authors translated the initiation codon GTG for residue 1 as Val
R:Korblum, J.; Projan, S.J.; Kreiswirth, B.N.; Mogazeh, S.L.; Elsner, W.; Ross, H.;
submitted to the EMBL Data Library, March 1990
A:Reference number: S20793
A:Accession: S20797
A:Molecule type: DNA
A:Residues: 1-83, 'T', 85-423 <KOR>
A:Cross-references: EMBL:X52543; NID:946505; PIDN:CAA36783.1; PID:g581546
C:Genetics:
A:Gene: agc
A:Start codon: GTG

Query Match 17.48; Score 379; DB 2; Length 423;
Best Local Similarity 21.98; Pred. No. 6.4e-16;
Matches 96; Conservative 107; Mismatches 163; Indels 72; Gaps 7;
OY 33 LSKKELTESISNPLIMAVTMVNNVLEFPAPLYFIALSIYLNQNSLSNIFYGLPLV 92
Db 4 LSNYFVFLVLOMLTMTIPATISGIVSKLDYFII-----IYI 43
OY 93 ASSDLFRRAIIFLIDGTQIGVMSIITTYMIEFAGIAL-----SYLFSLV 139
Db 44 STLSLF---LFKMPSASLILTSFIIYFVIXKYYSILLMSQIILYCANYWYIYI 99
OY 140 FVNDIGRLKDSITKMKVKRKLIPNITMMLLYLLIOVLYESVNYIPLTKPR--PVYIV 198
Db 100 Y-AYITKISDSIFVIFPSEFVYVYIISLFSYIINRYLAKIST---PYILNKGLIYI 154
OY 199 YLILFL-----LISPLSOYTKQVONEIMAKQEQ 229
Db 155 STILTLTFSLFFFYQSINDEKAVIROYSLFYWYHYLSITLYSOFLKEMKVKRNOEE 214
OY 230 INRITQYSQISLTKDRSRPHNYNLTSLRLGLENKPLASIEKTYHOILEKTHQLO 289
Db 215 IETYYEYITKIAINNEKRRHHDYVNLITLSEYIKEDDMPGIRDFYFNKNYIPMDNQ 274
OY 290 DTRVYNGHIANIQNDAVNGILSAILLEAONKKIAVNVSSKIQLEPMELDFITLSTL 349
Db 275 MNAIKLNGENLKVREIKGLITAKILRAOENNIPISEIDEVSSINLNMIDLSRSIGII 334
OY 350 CDNAIEAFESINPEIOLAEPFKNGSIYFIIONSTKQIDVSKIREKENTSTGSRGIG 409
Db 335 LDNAIATASIEIDPIIRAFIIESENSYFIYMNKCADIRIHELQVDESGTKGEGRGIG 394
OY 410 LAKVNHILEHPKTSLOQ 427
Db 395 LSTLKEIADNADNVLDT 412
RESULT 9
D89995
accessory gene regulator C [imported] - *Staphylococcus aureus* (strain N315)
C:Species: *Staphylococcus aureus*
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: D89995

R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguma, A.; Mizutani-U, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shibata, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratsuka, K.
Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
A:Reference number: A89758; MUID:21311952; PMID:11418146

A:Accession: D89995

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-371 <OR>

A:Cross-references: GB:BA000018; PID:g13701832; PIDN:BA843125.1; GSPDB:GN00149

A:Experimental source: strain N315

C:Genetics:

A:Gene: agrC

Query Match 17.1%; Score 372; DB 2; Length 371;

Best Local Similarity 23.5%; Pred. No. 1.3e-15;

Matches 101; Conservative 92; Mismatches 162; Indels 74; Gaps 8;

14 TYLVFLFLFLFSKVSNTLSKKELTFSISNFMIAVTWVNM.F---YPAEPLVF 68

4 TESSIFLEFKLYVAIVT-----LVMTIMVLSNFTVGLFLTKYTPDA-- 52

69 IASTIYINRONSLSNIFGGLPVASDFFRAIIFPILDTGQIVMGSSITTYMTEFA 128

53 ILPLVILSSVSLATY--LVKISLKKFKSYL-----SLNKTYMIIIS 96

129 GIAL-SYLVSPNDIGRLKDSLTKMKVKRLIPMNITMLLYLLIOVLVIESYVNP 187

97 FVLEAFAPFYISTWTSSNGDS-----LIPYALVIGLIIIFISVILIMSLFLK 147

188 TLKFRKRVVIVLLEFLILISFSQYTKQVONELMAQKQKQINPTQSOQISLKYDI 247

148 EMKYR-----NQEIEIYYETIKIEINNM 175

248 RSPRHLYNLITSLRGLENKDLASIEKTYHQLEKTHQLODTRYNIGHLANTIONDAV 307

176 RFRPHDVNLTLLSEIREDMDIGLRAVENKNIYPMKNDLQMAIKNGIENLKVREIK 235

308 GILSAKLEKONKIAVNEVSSKIQLPEMELDTITLISICDNAAIFAPESLPEIOL 367

236 GLITAKILRQENMIPISIEIPREVSSIMLMIDLSRSGITLLDNAIFASTEIDPIIRV 295

368 AFFKNGSIYFIIONSTKEKQDVSKIFENKSTGSGNFGIGLAKVNHILEHPTSLQT 427

296 AFIESENSYFIYMNKCADIPRIHELFOESFSTGEGRGLOLSTLKEIADNADVILDT 355

428 SNHHLEFKQ 436

356 IIEGFFIQ 364

RESULT 10
G56273
protein-histidine kinase homolog spk - *Lactobacillus sake* (strain Lb706) plasmid pLSA60

C:Species: *Lactobacillus sake*

C:Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 15-Oct-1999

C:Accession: G56273

R:Axelsson, L.; Holck, A.

J. Bacteriol. 177, 2125-2137, 1995.

A:Title: The genes involved in production of and immunity to sakacin A, a bacteriocin fr

A:Reference number: A56273; MUID:95238285

A:Accession: G56273

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-428 <AXE>

A:Cross-references: GB:Z46867; NID:9598449; PIDN:CAA86944.1; PID:g59854

C:Genetics:

A:Gene: spkK, spkB

A:Genome: Plasmid

A:Start codon: GTC

Query Match 16.7%; Score 362; DB 2; Length 428;
Best Local Similarity 24.8%; Pred. No. 7e-15;
Matches 111; Conservative 84; Mismatches 178; Indels 74; Gaps 12;

12 LTVLVFLFLFLFSKVSNTLSKKELTFSISNFMIAVTWVNM.FYPA 63

39 IAVFPIITVVALYSDISNVIPILSGVILKEKENDYLLIDLFCFTIFECNV----- 94

64 EPLVFLAISYINRONSLSNIFGGLPVASDFFRAIIFPILDTGQIVMGSSITTYM 123

95 -----LSSNIMQLIP-----SNRTVGF-----GIFQL 119

124 MIEFAGIALSYFLSVFVNDIGRLKDSLTKMKVKRLIPMNITMLLYLLIOVL---YVI 180

120 FIEALYISITIFPKKNH-----NVKESKASTVSLIYLLVLLISTYA 168

181 ESYNVIPLTKREV--VIYVIL---FLIISFLSQYTKQVONELMAQKQKQINPTQ 235

169 HYDA-----YDHFVLGIMIFLIQVFEVET--FLRMILTKQRTKKYKQLEKQ--ELNNLKK 221

236 YSOQIESLYKDIRSRHDYLNITSLRGLENKDLASIEKTYHQLEKTHQLODTRYN 295

222 YTESLEQOOQOQSKRPHRYKNLLSFKENITFNKKTALQIELEQYSNRYLDKGEFY 281

296 GHLANTIONDAVGLISAKILEKONKIAVNEVSSKIQLPEMELDTITLISICDN 355

282 KALYNHNEFVSLITAKIHQAKELNIECYECCKRPDIYPIPIFDIRILGILDNAIE 341

356 AAFESINPEIOLAFKNGSIYFIIONSTKEKQDVSKIFENKSTGSGNFGIGLAKV 415

342 AASECKELIYVYQDDQLEFISIKNTYKSNNSISTQLQKRNISTKKGSGGLWTIE 401

416 ILEHPTSLQTSNHHLEFK--QLIIR 441

402 ENOKFPNVFTQYKQESFESVOLLIR 428

RESULT 11
S57909
probable histidine protein kinase spk - *Lactobacillus sake*

C:Species: *Lactobacillus sake*

C:Date: 24-Mar-1997 #sequence_revision 24-Mar-1997 #text_change 15-Oct-1999

C:Accession: S57909

R:Huehne, K.; Holck, A.; Axelsson, L.; Kroeckel, L.

submitted to the EMBL Data Library, February 1995

A:Description: Cloning and nucleotide sequence of genes involved in regulation and se

A:Reference number: S57909

A:Accession: S57909

A:Molecule type: DNA

A:Residues: 1-448 <HUP>

A:Cross-references: EMBL:Z48542; NID:9695615; PIDN:CAA8426.1; PID:9695616

A:Experimental source: strain Lb674

C:Genetics:

A:Gene: spkK

Query Match 16.0%; Score 348; DB 2; Length 448;
Best Local Similarity 26.3%; Pred. No. 5.3e-14;

Matches 121; Conservative 92; Mismatches 199; Indels 46; Gaps 15;

8 LNSGLITVYLVFLFLFSKVSNTLSKKELTFSISNFMIAVTWVNM.FYPAEPLY 67

6 VSVSLMONFVALILFLLYRYIORKIFPKRIID-----ILAIIFSIYILPISDASIL 59

68 FIAL-----SIYLRON-----SLSNIFYGILPVASDFFRAIIFPI-LDGTG 113

60 VAVIMKLSGHHFQCKENIKITPDANLLIVLQDLVAVGIIISQTTISIKSDSQNI 119

114 VNGSITTYMIEFAGIALSYFLSVFVNDIGRLKDSLTKMK--VKRRLIPMNITMLLY 172

120 LNSNDITILL-----GIFFAVLFDGLFPI--LKNKRTTELQHLNDIEFISDEK--QYFI 171

QY 173 LIOVLT-VIESNVIPETK-FREVVIVYLLLELILIS-----FLSOYT-KQKVONE 221
 Db 172 FIFILFIVEILLAVGNLQGVATALLIIFLCVLLGMFWOMVFLAKYSROEDNQ 231
 QY 222 IMAKQEAQIRNITQYSSQOIESLYKDIRSFHDYDNIITSLRLGIENKDLASIKIHOITL 281
 Db 232 LVANQOQL-----DVLVINEQGYTELRFKNDYQNTLLISLESAEKGDQOQFAYVQELL 286
 QY 282 EKHGHOLOQTRYNIGHLANIQNDVAKGILSKILEAQNKIAVNVESKIQLEPEMLLD 341
 Db 287 AQRPQISEIQGAVIAQLOLYLKNQDPRIGLIVLOKFLAAQAVTLKFEPTPEIELATANLLT 346
 QY 342 FIFILSLCDNAIEAFESLNPETQLAEFFKNGSIVFIIONSTREKQIDVSKITFE-NYS 400
 Db 347 VIRIIGIILDMNIEQAVQETDQVSCAFLOSGDLIEITIENTA--SQVKNLQAFSELGYS 404
 QY 401 TKSNGNGIGLAKVNHILEHYKRTSLQTSNNHHLFKOLLIT 440
 Db 405 TKAGRGKGLANVQDLIAKQTNLPLETOIENRKLQRTLMT 444

RESULT 12

B97095

membrane associated histidine kinase-like Atkase [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum
C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001

C:Accession: B97095

R:Noiling, J.; Breton, G.; Omechenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: B97095

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-452 <KUR>

A:Cross-References: GB:AE001437; PIDN:AAK79549.1; PID:g15024536; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

A:Genetics:

A:Gene: CAC1582

Query Match 10.3% Score 223.5; DB 2; Length 452;
 Best Local Similarity 21.6% Pred. No. 2e-06;
 Matches 102; Conservative 97; Mismatches 186; Indels 87; Gaps 18;

QY 1 MNEALMISNGILYLYLFLFLFESKVSNTLSKKELEFSSNPLMIATVNVNLF 60
 Db 1 MNE-IQIYINSEITNALIAIKYFYREBELGFSKRRKIYLFVLEIIVSISNVSCKLG 59
 QY 61 YPAEPLXYF--ALSTYLRONSLSLNFYGLPVASDLEFRRALIFPILIDGTQIVMSS 118
 Db 60 FKSILIIILIEIILIFLCRGSKSLVKIYSVL--TENILILSLTFLIYDPRKIDHYSRA 117
 QY 119 IYTYWIEFAGIALSYFLSVFNVNIGRLKDSLTK-----MKVKKRLPMTNI 165
 Db 118 FINKD-----YKIFENLILNFRDILNIEFLMGLKICISIIKINS--ELNI 163
 QY 166 TMLLYYL-----IOVLVYESVNV-----IFTLKRFVVIYVLLIFELILI 207
 Db 164 YKQLYIVMPCISYIGISLF-----YVQVLDYTGGEYLEIFESSIYCMLEPIGLIILIS 219
 QY 208 SFLSQYTKQV-----QNEIMAKQEAQIRNITQYSSQOIESLYKDIRSFHDYDNIITLS 260
 Db 220 IFIAAATFOKMLDGEEREKESILMOQOQNLQ--FEHKKRNQNTYKNTRGVMDINKHINC 277
 QY 261 LRLGLENKDLASIEKIQILEKTHQLODPRYNIGHLANIQNDVAKGILSKILEAQNK 320
 Db 278 LR-----NLAYSQNI--DVLKNTLNNISEIKKLDPILITGAVAADVAINKTYITAKNE 329
 QY 321 KIAVNVESKIQLEP--MELDPITILSLCDNAIEAF-----SLNPEQLAEFFKN 373
 Db 330 ---GIELYCFIMEPEKAGIEPIDLSIILNNSLDNAIEACRKIKNLKRLISIKSIYKN 385

QY 374 GSIFLIIONSTREKQIDVSKITFENYSTKGS--NRGIGLAKVNHILEHYKPT 423
 Db 386 SFELIIEISNCEE--GIKYCNNNNISTKSETHGIGISNIKEAVKYSYGT 434

RESULT 13

AB1533

two-component sensor histidine kinases homolog lln0802 [imported] - Listeria innocua

C:Species: Listeria innocua

C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001

C:Accession: AB1533

R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec

; Dominguez-Bernal, G.; Duclaud, E.; Durand, L.; Dussurget, O.; Entlian, K.D.; Fsiht,

D.; Jones, L.M.; Kerst, U.

Science 294, 849-852, 2001

A:Authors: Krefte, J.; Kunz, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;

Ok, C.; Schueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla

A:Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AB1533

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-433 <GLA>

A:Cross-References: GB:AL592022; PIDN:CAC96034.1; PID:g16413253; GSPDB:GN00178

A:Experimental source: strain C1p11262

A:Genetics:

A:Gene: lln0802

Query Match 8.5% Score 184.5; DB 2; Length 433;
 Best Local Similarity 24.1% Pred. No. 0.00044;
 Matches 110; Conservative 79; Mismatches 155; Indels 113; Gaps 25;

QY 16 LVYVLPFLPFS--KVSNTLSKKELEL-FSISNPLMIATVNVNLFPAEPLXFIALS 72
 Db 17 LEVLLIYFESKRYKMPNPRVSKERLLIFCIASL-----VOAFDSFE-YF----- 61
 QY 73 IYLRONSLSLNFYGLPVASDLEFRRALIFPILIDGTQIVMSSIYTYMIEFA-GIA 131
 Db 62 -----TINV-----IVSAVLFL-VVSFEFFKQODILM-SIITATVAIDIGEGIA 104
 QY 132 LSVLPSVNVNDIGRLKSLTKMKVKKRLIPMTIMLYL-----LIOVLVIESVNV- 185
 Db 105 L-----MSLSTALIGR-----QDVGEMLQPSNTMLFLISRLISLGLVLLILFRKIR 153
 QY 186 -----IPTLKERFVVIY--LILFLL-----ISF-----LSQYTRQK 217
 Db 154 LGNSLPTIYVTFPLFTMKDILMWLLIADSPVEISFYWAAVIPLPVSYILFYTRRS 213
 QY 218 YONEIMAKQEAQIRNITQYSSQOIESLYKDIRSFHDYDNIITSLRLGIENKDLAS 272
 Db 214 IEMVATQVDSVLDDEKKKYTEQDLVTKQLESQRTVRHDLKKNISPL--TYLAESGK 270
 QY 273 IEKIQHILEKTHQLODPRYNIGHL-----ANTQNDVAKGILSKILEAQNKIAVN 325
 Db 271 TELVEIQVOE-----LGSLSVLGLIYADSGNITIDYIILKQALANRGVYIF 318
 QY 326 VEVSSKIQLEPEMLDPITILSLCDNAIEAFESLNP-EIOLAEFKKGSIVFIIONST 384
 Db 319 CEINVPNDI-DVPPDLSTVLGNLIDNAARLEEVONEKRTIGIKISYQVGFMLIOVANSF 377
 QY 385 KE-KQIDVSKITFENYSTKGSNRGIGLAKVNHILEHY 420
 Db 378 DGIIVLDNKKIISREKDE--NHGIGLISIKIKIAGY 412

RESULT 14

E69600

two-component sensor histidine kinase cIts - Bacillus subtilis

C:Species: Bacillus subtilis

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000

C:Accession: E69600

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bered
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallen
leeh, J.; Harwood, C.R.; Henauf, A.; Hilbert, H.; Holappel, S.; Hosono, S.; Hullo, M.F.
Koeber, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y, M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sabate, Y.; Sato, I.; Scanlon,
A.; Schleich, S.; Schroeder, R.; Scottone, F.; Sekiguchi, J.; Sekowska, A.; Seton
deuchl, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tononhi, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wypat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
A:Reference number: A69580; MUID:98044033
A:Accession: E69600
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-542 <KUN>
A:Cross-References: GB:299108; GB:AL009126; NID:92633055; PIDN:CA12587.1; PID:92633082
A:Experimental source: strain 168
C:Genetics:
A:Gene: cits
C:Superfamily: two-component sensor histidine kinase; sensor histidine kinase homology

Query Match 8.0%; Score 174; DB 1; Length 542;
Best Local Similarity 24.1%; Pred. No. 0.0025;
Matches 97; Conservative 73; Mismatches 134; Indels 98; Gaps 21;
QY 35 KKELEFSSINFLIMAVTVVNVNLFYPAEPFLFALSTYLN--RONSLSNIFGLPV 92
DB 171 KKHRLSLSTYATLVLL-----LGFIGNAVLAKSKRKOTL-----GLEPH 209
QY 93 ASSDLF--RAIIFILDGOGIVNGSSITFTYMIFFAGIALSYLFLSVFNVDIGRLKDS 150
DB 210 EIALALYRERNAMLFAIRG-----LIAT--NREGV-----VTMMVSAEM--- 248
QY 151 LFKMKKKRLIMNTMLLYLLIOVLVIESYVITPLKPKFVIVYLLPLILISFL 210
DB 249 ---LKLPEVHILPIDDVMPG--AGIMSYLKEGEMLPNOE-----VSYNDOVFII----- 293
QY 211 SOTTKOKVONE-----IMAKQEAQIRNITQYSQIESLYKDIRSRHDYLN---IL 258
DB 294 ---NKKVNNOCGAYGIVVSFRKTELKLLIDLEVRKYSDDLRAOTHEFSNKLAYAIL 349
QY 259 TSLRLGIEKKDLASIEKIYHQLEKTHQLODTRNIGHLANIONDAVKGLSAKILEAQ 318
DB 350 GLELGEYDEALDIKEEY--AIQNEQHDL-----LFH--NIHSQOVQAILLGLKISKAS 399
QY 319 NKKIANNV-EVSSKITQLP-EMELLDFTILSLCDNAIEAFAESINPEIOLAEPKNGST 376
DB 400 EKKVKLVIDENSLAPLPAHIGLSHLITLIGNLIDNFEAFAEQSVKEV-LFFITDMGHD 458
QY 377 VFIIQNST-----KEKQIDVSKIFKENSSTGKSNNGIGLAKV 413
DB 459 IYIEVSDTGPQVPPK---IEAVFERGYSKGMRRGIGLANV 497

RESULT 15
E83699
two-component sensor histidine kinase BH0397 [Imported] - *Bacillus halodurans* (strain C-
C:Species: *Bacillus halodurans*
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_Change 15-Jun-2001
C:Accession: E83699
R:Itakami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirai
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: E83699
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-532 <STO>
A:Cross-References: GB:AF001508; GB:BA000004; NID:910172890; PIDN:BA004116.1; GSPDB:G
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH0397
C:Superfamily: two-component sensor histidine kinase; sensor histidine kinase homology

Query Match 7.9%; Score 171.5; DB 2; Length 532;
Best Local Similarity 24.0%; Pred. No. 0.0035;
Matches 53; Conservative 44; Mismatches 87; Indels 37; Gaps 5;
QY 218 YQNEIMAKQEAQIRNITQYSQIESL-----YDIRSFRHDYLNITSLRLGIEKKDLA 271
DB 299 IYNDVVGAIISFRDLFEIKQQAQOLGVKLYEALQAQSHPEFNKLHVILGWKTESYD 358
QY 272 SIEKTYHQLEKTHQLODTRNIGHLANIONDAVKGLSAKILEAQNKKIANNVEVSSK 331
DB 359 ELNDYTHQIVNHRSTELNHV-----IKRKDSVLAQFIIKLSYAREKHITLDVQTKSV 412
QY 332 IQLPEMELL--DFTILSLCDNAIEAFAESINPEIOLAEPKNGSIVFTIQNSTREKQI 389
DB 413 IPKASSEQVHVELITLIGNLIDNLEAVIKA-----KEKTYLVQLEYSNRLHI 461
QY 390 DY-----SKTKENYSTKSGNRGIGLAKVNHILE 418
DB 462 RVQDTGPGIPDQGGIDPKKGYSTKENRGYGLYVQSVE 502

Search completed: July 30, 2002, 15:15:20
Job time: 391 sec

Wed Jul 31 08:30:58 2002

us-09-833-017-4.rpr

Page 8

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:20:52 ; Search time 15.86 Seconds
(Without alignments)
1076.628 Million cell updates/sec

Title: US-09-833-017-4
Perfect score: 2173
Sequence: 1 MNEALMILSLNGILTYLFLV.....KTSLSQTSNHHHLFKOLLIIK 441

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000.

Post-processing: Minimum Match 0%,
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	151	6.9	543	1	DCUS_ECOLI
2	139	6.4	428	1	KINB_BACSU
3	129.5	6.0	532	1	DCUS_BACSU
4	129	5.9	880	1	YE21_ARCTU
5	128.5	5.9	552	1	DP1B_ECOLI
6	127	5.8	577	1	CHVG_RHIME
7	126.5	5.8	535	1	DCUS_BACSU
8	126.5	5.8	691	1	LCN3_LACIA
9	125.5	5.8	717	1	NTP2_MSEPV
10	125.5	5.8	938	1	VL20_HSVTJ
11	120.5	5.5	430	1	SUPA_BUCAT
12	118	5.4	1024	1	RPOB_PLARA
13	117.5	5.4	214	1	SP2K_BACSU
14	117.5	5.4	388	1	YF02_AOUAE
15	116	5.3	656	1	YC26_PORPU
16	116	5.3	850	1	DP01_ANAHI
17	115.5	5.3	451	1	YG1Y_HAEIN
18	115.5	5.3	873	1	COX1_ACACA
19	114.5	5.3	600	1	NIST_LACIA
20	114.5	5.3	904	1	TIR3_HUMAN
21	113	5.2	366	1	Y976_METUA
22	112.5	5.2	382	1	Y061_MESVI
23	112.5	5.2	883	1	YH66_YEAST
24	112.5	5.2	1174	1	EX5B_BUCAT
25	112	5.2	496	1	YFHK_ECOLI
26	111.5	5.1	780	1	MUS2_BORBU
27	111	5.1	556	1	NU2M_PODAN
28	110.5	5.1	886	1	RA50_ARCTU
29	109.5	5.0	440	1	CAPE_STRAU
30	109.5	5.0	1274	1	BXF_CLOBO
31	109	5.0	1121	1	ALP1_SCHPO
32	108.5	5.0	455	1	TRME_LACIA
33	108.5	5.0	614	1	YDNK_LACIC

34	108	5.0	339	1	Y83C_METUA
35	108	5.0	608	1	ATOS_ECOLI
36	108	5.0	885	1	YDGH_BACSU
37	108	5.0	1036	1	YG55_YEAST
38	108	5.0	1738	1	YCF1_EPIVI
39	107.5	4.9	336	1	YPH1_MYCCA
40	107.5	4.9	490	1	PIT_BUCAT
41	107.5	4.9	2136	1	YCF2_MARPO
42	107	4.9	720	1	STFB_PORPU
43	107	4.9	844	1	BRO1_YEAST
44	107	4.9	1294	1	YA3B_SCHPO
45	106.5	4.9	660	1	Y390_MYCPN

ALIGNMENTS

RESULT 1
ID DCUS_ECOLI STANDARD; PRT: 543 AA.
AC P39272; P76795;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Sensor protein dcus (EC 2.7.3.-).
GN DCUS OR B4125 OR Z5727 OR ECS5107.
OS Escherichia coli, and
OC Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT *Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes.*;
RL Nucleic Acids Res. 23:2105-2119(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT *Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.*;
RL Nature 409:529-533(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi K., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe F.,
RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT *Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.*;
RL DNA Res. 8:11-22(2001).
RN [4]
RP CHARACTERIZATION.
RC STRAIN=K12 / MC4100;
RX MEDLINE=9840434; PubMed=9765574;
RA Zientz E., Bongaearts J., Unden G.;
RT *Fumarate regulation of gene expression in Escherichia coli by the
RT dcus (dcus genes) two-component regulatory system.*;
RL J. Bacteriol. 180:5421-5425(1998).
RN [5]
RP CHARACTERIZATION, AND TOPOLOGY.

CC STRAIN-K12 / MC4100;
 CC MEDLINE-99138745; PubMed-997351;
 RA Goldy P., Davies S., Kelly D.J., Guest J.R., Andrews S.C.;
 RT "Identification and characterization of a two-component sensor-kinase
 RT and response-regulator system (DcuS-DcuR) controlling gene expression
 RT in response to C4-dicarboxylates in *Escherichia coli*.";
 RL J. Bacteriol. 181:1238-1248 (1999).
 CC -1- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM DCU/DCUS.
 CC INVOLVED IN THE C4-DICARBOXYLATE-STIMULATED REGULATION OF THE
 CC GENES ENCODING THE ANAEROBIC FUMARATE RESPIRATORY SYSTEM (FEDABCD;
 CC NUCON; DCUB; DCUC; SDHDA; ETC.). WEAKLY REGULATES THE AEROBIC
 CC C4-DICARBOXYLATE TRANSPORTER DCTA. PROBABLY ACTIVATES DCUR BY
 CC PHOSPHORYLATION.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
 CC -1- MISCELLANEOUS: THE REGION ENCOMPASSING APPROXIMATELY RESIDUES 42
 CC TO 181 HAS BEEN SHOWN TO BE PERIPLASMIC, HOWEVER EXACTLY WHICH
 CC RESIDUES ARE PERIPLASMIC IS NOT CLEAR.
 CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U14003; AAA97025.1; -;
 DR EMBL: AE000485; AAC77086.1; -;
 DR EMBL: AE005646; AAG59324.1; -;
 DR EMBL: AP002568; BAB38530.1; -;
 DR EcGene; EG12465; dcus.
 DR InterPro; IPR004358; BCTRLSENSOR.
 DR InterPro; IPR003594; HATPase_C.
 DR InterPro; IPR000014; PAS_kin_sig.
 DR InterPro; IPR000014; PAS_kin_sig.
 DR Pfam; PF02518; HATPase_C; 1.
 DR Pfam; PF00989; PAS; 1.
 DR Pfam; PF00512; signal; 1.
 DR PRINTS; PR00344; BCTRLSENSOR.
 DR SMART; SM00387; HATPase_C; 1.
 DR SMART; SM00091; PAS; 1.
 DR PROSITE; PS50109; HIS_KIN; 1.
 DR PROSITE; PS50112; PAS; 1.
 KM Sensory transduction; Transferase; Kinase; Transmembrane;
 KM Inner membrane; Phosphorylation; Complete proteome.
 FT DOMAIN 1 20
 FT TRANSMEM 21 41
 FT TRANSMEM 42 181
 FT TRANSMEM 182 202
 FT DOMAIN 203 543
 FT DOMAIN 220 291
 FT DOMAIN 346 538
 FT MOD_RES 349 349 HISTIDINE KINASE.
 FT MOD_RES 349 349 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 SQ SEQUENCE 543 AA; 60550 MW; 653C369344CBA238 CRC64;

Query Match 6.9%; Score 151; DB 1; Length 543;
 Best Local Similarity 23.4%; Pred. No. 0.041;
 Matches 89; Conservative 68; Mismatches 143; Indels 80; Gaps 18;

QY 85 IFGLLPVASSDLE--RAAIFIFIDGOGIVMGSSITTYMIFAFALSYL-----F 136
 DB 208 ILEGLPEYELSTLEFQRQMLDSIKGYVAVDDGGEV--TLINAAQELNLYKRSQDDER 265
 QY 137 LSVFN-----VDIGR-LKDSLTK---MKVKKRLIPNITMLLYLLIOVLYIESYAV 185
 DB 266 LSTLSHWSQVVDSEVLDTGPRRDEITIKDRLLIN----- 304
 QY 186 IPTLKFRKFFVYIYLLIFLISPSOYTKQVONEIMAKQEAQIRNITQYQSIESLTK 245
 DB 305 --IVPVRKNGVT-----IGALSTP-----RDKTEVRKLMGRDLGLVAVNYAD 342

QY 246 DIRSFHDYINTLTSLRIGENKDLASTEKIYHIIKTEKTHGLODTRVNGH-LANIOND 304
 DB 343 ALNERSHEEMKRLHVI-LGLHL--LKSTKQLEDTYL-KTANNQOE--ELGSLGKIKSP 395
 QY 305 AVKGLSARKLEAOKKIAVNEVSSKIQLEPMELDFITILSDMAIEAFESLND- 363
 DB 396 VIAGFLISKINRATP--LGHVLLINSESQLDPSGSEQVATLLTTLGNIEVALGPE 453
 QY 364 ---EIOIAFFKNGSIVFLIONSTREKQID-VSKTFEKNYSTKNGSLAKVNHLEH 419
 DB 454 PGCEISVTLHYHNGHMLHCEVNDGPGIAPKIDHIFPKGVSTGSRGVALVQGVEN 513
 QY 420 YPKTSIQTSNHHHLKQLII 439
 DB 514 L-GGSIAVESEPGIFTOFFV 532

RESULT 2
 ID KINB_BACSU STANDARD; PRT; 428 AA.
 AC 008450;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Sporulation kinase B (EC 2.7.3.-).
 GN KINB.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxId=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RX MEDLINE-93268102; PubMed-8497199;
 RA Trach K.A., Hoch J.A.;
 RT "Multisensory activation of the phosphorylation initiating sporulation
 RT in *Bacillus subtilis*: identification and sequence of the protein
 RT kinase of the alternate pathway.";
 RL Mol. Microbiol. 8:69-79 (1993).
 RN [2]
 RP REVISIONS.
 RC STRAIN-168;
 RA Darbois V.A., Djavahishvili T., Hoch J.A.;
 RL Submitted (Aug-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RX MEDLINE-97419515; PubMed-9274030;
 RA Oudega B., Koningssteijn G., Rodrigues L., de Sales Ramon M.,
 RA Hilbert H., Duesterhoeft A., Pohl T.M., Weitzengger T.;
 RT "Analysis of the *Bacillus subtilis* genome: cloning and nucleotide
 RT sequence of a 62 kb region between 275 degrees (rmb) and 284 degrees
 RT (pal).";
 RL Microbiology 143:2769-2774 (1997).
 CC -1- FUNCTION: PHOSPHORYLATES THE SPOULATION-REGULATORY PROTEINS
 CC SPOOA AND SPOOF. SPOOF IS REQUIRED FOR THE KINB ACTIVITY.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z93933; CAB07911.1; -;
 DR EMBL: U63302; AAB61980.1; -;
 DR EMBL: Z99120; CAB15134.1; -;
 DR PIR: S32935; S32935.
 DR Subtilist; Bg10745; kinB.

DR InterPro: IPR003594; HATPase_C.
 DR InterPro: IPR004359; HIS_KIN_S19.
 DR InterPro: IPR003661; HIS_KIN_S19.
 DR Pfam: PF02518; HATPase_C; 1.
 DR Pfam: PF00512; signal; 1.
 DR SMART: SM00387; HATPase_C; 1.
 DR SMART: SM00388; HSKA; 1.
 DR PROSITE: PS50109; HIS_KIN; 1.
 KM Sensory transduction; Transferase; Kinase; Phosphorylation;
 KW Sporulation; Transmembrane; Complete proteome.
 FT DOMAIN 1
 FT TRANSMEM 7 27
 FT DOMAIN 28 37
 FT TRANSMEM 38 57
 FT TRANSMEM 58 68
 FT TRANSMEM 69 89
 FT TRANSMEM 90 99
 FT TRANSMEM 100 120
 FT TRANSMEM 121 132
 FT TRANSMEM 133 153
 FT TRANSMEM 154 166
 FT TRANSMEM 167 187
 FT TRANSMEM 188 428
 FT DOMAIN 218 221
 FT MOD_RES 221 426
 FT CONFLICT 44 48
 FT CONFLICT 328 48
 SQ SEQUENCE 428 AA; 47812 MW; 52FBECA6864F7040 CRC64;

Query Match
 Best Local Similarity 20.5%; Score 139; DB 1; Length 428;
 Matches 93; Conservative 86; Mismatches 188; Indels 86; Gaps 17;

QY 7 ILNGLITLVYVLLPLFLFSKSVNTLSKELTLESINFLIMAVMNV-NLFYAPR 65
 3 ILKDYLLHICFLIFLLYQV--FWLGPALVVKINSSGLVTLFACGASVLCIIFPIHE 59
 DB 66 LYLIALSIYLNQNSLSLNIIFYGL--LPAVSSDLFRRALIFFIIDGQIVMGSSITTY 123
 QY 60 MDYIQ-----YGLQMIPV-----ITCLFYISTASGLFVAASVCFE 95
 DB 124 MIEFALISYLFSLFENVDIR--KOSLYMKYKRLIMNTMLLYLLQVLYIE 181
 DB 96 LLETPSAFVFTLPELIIIPILQKMPKSKAKLLLSLSCVEIFLFPASSWILS 155
 QY 182 SYNVIPFLKFRKVVYVLI---LELLISFLSQYKQKQVNEIMAKQKQINRITQYS 237
 DB 156 ALNI---LNFQKSGIFVEYEAASGLFRSSVLLSIYIIESIAENI-----ALRSQLIHS 206
 QY 238 QOIESLYKDIRSRHDYLLSLR-----LGTINKDLASIEKTYHQILEKTGHOL 288
 DB 207 EKMTIYSELAAVAHVRPLTVVGRFVOLFNEDTLDKSSADYKLVLSLDRAO--- 263
 QY 289 QDTRNYIGHLANIQNDVAVGI---LSAKILEAQKKI-----AVNVESSKIQLP 335
 DB 264 -----GIITNYLDMARKQOLEKEVEFDLSALIKETSLMAYNKSVTEATE--P 313
 QY 336 EMELIDFTILSLCDNATEAFESL--NPEIOLAFKKKSGSYF--IIONSKRQID 390
 DB 314 DLLIYGDARKLKQAVINLMKNSIEAVPRCKGMHIS-AKRNGHITMINITNGVGMTHQ 372
 QY 391 VSKIKENSTYKSGNRGICGLAVNHLIEHPKT 423
 DB 373 MOKLGEPPYSLKNTGTLGLTVTFSLIEHHNGT 405

RESULT 3
 DCTS_BACHD STANDARD; PRT; 532 AA.
 AC Q9K957;
 DT 16-OCT-2001 (rel. 40; Created)
 DT 16-OCT-2001 (rel. 40; Last sequence update)

DT 01-MAR-2002 (rel. 41; Last annotation update)
 DE Probable C4-dicarboxylate sensor kinase (EC 2.7.3.-).
 GN DCTS OR BH2752.
 OS Bacillus halodurans.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/staphylococcus group; Bacillus.
 OX NCBI_TaxID=86665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C-125 / JCM 9153;
 RX MEDLINE=20512582; PubMed=11058132;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
 RA Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
 halodurans and genomic sequence comparison with Bacillus subtilis";
 RL Nucleic Acids Res. 28:4337-4331(2000).
 CC -I- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM DCTS/DCTR.
 CC PROBABLY ACTIVATES DCTR BY PHOSPHORYLATION. ESSENTIAL FOR
 CC EXPRESSION OF DCTR (BY SIMILARITY).
 CC -I- SUBCELLULAR LOCATION: Integral membrane protein (potential).
 CC -I- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
 CC -I- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AP001516; BAB06471.1;
 DR InterPro: IPR004358; BCTRSENSOR.
 DR InterPro: IPR003594; HATPase_C.
 DR InterPro: IPR004359; HIS_KIN_S19.
 DR InterPro: IPR000014; PAS.
 DR Pfam: PF02518; HATPase_C; 1.
 DR PRINTS: PR00344; BCTRSENSOR.
 DR SMART: SM00387; HATPase_C; 1.
 DR SMART: SM00091; PAS; 1.
 DR PROSITE: PS50109; HIS_KIN; 1.
 DR PROSITE: PS50112; PAS; FALSE_NEG.
 KM Sensory transduction; Transferase; Kinase; Transmembrane;
 KW Phosphorylation; Complete proteome.
 FT DOMAIN 1 12
 FT TRANSMEM 13 33
 FT TRANSMEM 34 175
 FT TRANSMEM 176 196
 FT TRANSMEM 197 532
 FT TRANSMEM 216 279
 FT DOMAIN 315 531
 FT MOD_RES 339 339
 FT SEQUENCE 532 AA; 59999 MW; 5C2445CEA0119A12 CRC64;

Query Match
 Best Local Similarity 18.4%; Score 129.5; DB 1; Length 532;
 Matches 97; Conservative 99; Mismatches 173; Indels 159; Gaps 22;

QY 4 AMITSNGLITLVYVLLPLFLFSKSVNTLSKELT----- 39
 DB 24 ALMWVSISLGLGVSI-----KEDELSNRTMTIQLVAQNHTVOQWDAPKPEASRTLOPI 79
 QY 40 ---LFSISNFLIMAVMNVNLFY-----AEPFYIALSIYLNQNSLSL 83
 DB 80 VERIRVINDHYIVLLNMDRIRITHPIPERLQPFVGVGDEDPAPF--AEHLYSKAKTEGV 137
 QY 84 NIFYGLPVASSDLFRRALIFFIIDGQIVMGSSITTY--MIE----- 126
 DB 138 VTVRAFPFLMQO--KEQV-----GVAVGVSLPSTADMTQEFWQPALIGLITAL 186
 QY 127 --FAG--IALSYLFLSYFNVNDIGRLKDSLTK----- 153

```

Db 187 PGTWGSMLASHIKRQTFNMEPDLAHLVPRDASFAHIEGVAINKHKEKTIINMEAR 246
Oy 154 --MKVKKRLIPMNTMLLYLLIQLVYISVNIPTLKRFRKRVV--VYLLIFLI--- 205
Db 247 RMLGVKRAIGRNI-----HEVTPDKLPJELLISIGKPKLYQREFFYIOGR 289
Oy 206 LISFISQYTKQVQNEIWA-----OKEAQIRNIQYQSQIESYKIDRFRHLYNLMS 260
Db 290 LV--FSNKPILQIDGEIYGALAIPODKSDVRLAEELTGVAEPDALRVONHEYSKLIH 347
Oy 261 ---LRIGIENKDLASIEKRYHQLEKTHGLODTRRNIGHLANIONDAVKGLISAKILE 316
Db 348 IAGLQIDEGKKALQYIDLEEEQEFSGVYMQ-----KIHNDISLAGLLGKFRK 397
Oy 317 AQNKRIAVNVEYSK-IQLPE-MELLDPITTLISCDNATEA--AFESLNPEIQLAFKK 372
Db 398 GKELQVOYIIEKDESEFIHPEGVTHDLVYVGMIDNSLDASISQODQKKYHVFIGE 457
Oy 373 NGSIVFIIIONSTREKQIDV-SKIFENYSTKG-SNRGIGLAKVNHILE 418
Db 458 NDFLKIRVADNGEGIREVERKEMVRGPFSTKSTSGRIGLFLIOAIVE 505

RESULT 4
YE21_ARCFU STANDARD: PRT: 880 AA.
ID YE21_ARCFU
AC 028851;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
GN Hypothetical protein AFI421.
OS Archaeoglobus fulgidus.
OC Archaea: Euryarchaeota: Archaeoglobales: Archaeoglobaceae:
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=9804943; PubMed=9389475;
RA Kleink H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrle N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirsnes E.F., Dougherty B.A., McKenney K., Adams M.D., Loftis B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -I- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -I- SIMILARITY: BELONGS TO THE UPF0182 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: AE001005; AAB89825.1; -
DR TIGR: AFI421; -
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 19 41 POTENTIAL.
FT TRANSMEM 61 83 POTENTIAL.
FT TRANSMEM 103 125 POTENTIAL.
FT TRANSMEM 152 174 POTENTIAL.

```

```

FT TRANSMEM 195 217 POTENTIAL.
FT TRANSMEM 237 259 POTENTIAL.
FT TRANSMEM 268 290 POTENTIAL.
SO SEQUENCE 880 AA; 102210 MW; 8AD99AB74E50671 CRC64;

Query Match
Best local similarity 20.7%; Pred. No. 1.4;
Matches 113; Conservative 82; Mismatches 177; Indels 174; Gaps 27;

Oy 10 NGLLTYLT-VLEFLFESKYSNNTLSKELTFLSISNELLIMAVVNVNLTPEAPPL-- 66
Db 52 NVFLTYLKHSGFFFL-----EPITFVPLFLTNAAVR--KVLTFHGBPLKI 97
Oy 67 -----YFIALSIYLNQNSLSLNIFFGLPVAASDLFRRAILFTLDGT----- 110
Db 98 PHLADFGIAFFAALTV-----KNWSSMLFY-----FNSSNFVSDPIFGIDAAFTFOLP 148
Oy 111 -QGIYMGSSITTYMIEFAGIALSYFLSVFNVDIQLADSLTKMKVKKRLIP---MNI 165
Db 149 FLKIVIG-SLLAALLISLAIAPAYMTAF-----RWKSLDEE--KEIFPGSGFMHF 197
Oy 166 TMLLY--YLLIOLVLYIESYVNI-----PTLKE-----LAVNVE 191
Db 198 SALLFASPLLSAALILSRFEIVHSEGLISGASYVDVNILSPSLMELAILLSGIFAA 257
Oy 192 -----RKFFVIVVLLLELLISFLSOYTKOK--YONEIWAQKEADIRNTOYSGOI 240
Db 258 YIVARRVERVFQIVGVLFVATILTFVAPFVQKFEVPSSEISYGRINANSINTLFA 317
Oy 241 ESLYKDIRSPHDYLNLTLSRLGIENKDLAS-----IEKIHQILE-KTGHOLOD 290
Db 318 YNLH-DVKLOQFDYDWDVYVNDI-LEAKPTIDNIRIMDHPRIDVPRQLQIRTYVIOD 375
Oy 291 T---RYNIGHLANIONDA-VKGLISAKILEAONK-----LAVNVE 327
Db 376 VDVKRYRI-----NDISYQULLAARELSTENLPSSAQOTWLKHLITRHGVIYASPN 428
Oy 328 VSSKIQLPEMELDLFTTLISLIDCNATEAFAESLNPEIQLAFKKNGSIVFIIIONSTREK 387
Db 429 KYSKRGLEDEFYIYDI-----PPEGLIKIERPEIYVGGELTDYVVVNTLQNEEDFP 478
Oy 368 QIDVSKIRKENYKVR-GSNRGIGL-----AKVNHILEHYPKTSLQSNHHLF 434
Db 479 KGDV-----NFKTYAGDGKIKLDYRKVLFSPRFEDINLISNITPDESRLAMHRODI 532
Oy 435 KQLLTI 440
Db 533 ERVSTI 538

RESULT 5
DP1B_ECOLI STANDARD: PRT: 552 AA.
ID DP1B_ECOLI
AC P77510; 054338;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Sensor kinase dp1b (EC 2.7.3.-) (Sensor kinase cta).
GN DP1B OR CITA OR MPDB OR B0619.
OS Escherichia coli.
OC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=KL2 / MG1655;
RX MEDLINE=98367124; PubMed=9701802;
RA Ingner H., Miller C.A., Cohen S.N.;
RT "Destabilized inheritance of pSci01 and other Escherichia coli
RL plasmids by DpiA, a novel two-component system regulator.";
RN Mol. Microbiol. 29:49-59(1998).

```


RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE-97426617; PubMed-9728503;
 RA Blatner F.R., Plunkett G., Tili, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Valdes J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.,
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,
 RA Federspiel N., Hyman K., Kaiman S., Komp C., Kuri O., Lew H., Lin D.,
 RA Natch A., Oetner P., Roberts D., Schramm S., Davis R.W.,
 RL Submitted (Jan-1997) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE-97061202; PubMed-8905232;
 RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
 RA Sasaki G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiiuchi T.,
 RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
 corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-155(1996).
 CC -1- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM DP1A/DP1B
 CC ESSENTIAL FOR EXPRESSION OF CITRATE-SPECIFIC FERMENTATION GENES
 CC AND GENES INVOLVED IN PLASMID INHERITANCE. MAY ACTIVATE DP1A BY
 CC PHOSPHORYLATION.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (Potential).
 CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: U46667; AAC28951.1; -
 DR EMBL: AE000167; AAC73720.1; -
 DR EMBL: U82598; AAB40819.1; -
 DR EMBL: D90702; BAA35255.1; -
 DR EMBL: D90703; BAA35261.1; -
 DR Ecogene: EG13646; dplB
 DR InterPro: IPR004358; BCRRLSENSOR.
 DR InterPro: IPR003594; HATPase_LC.
 DR InterPro: IPR004359; HIS_KIN_S19.
 DR InterPro: IPR00014; PAS.
 DR Pfam: PF00518; HATPase_C; 1.
 DR Pfam: PF00989; PAS; 1.
 DR Pfam: PF00512; signal; 1.
 DR PRINTS: PR00344; BCRRLSENSOR.
 DR SMART: SM00387; HATPase_C; 1.
 DR SMART: SM00091; PAS; 1.
 DR PROSITE: PS50109; HIS_KIN; 1.
 DR PROSITE: PS50112; PAS; FALSE-NEG.
 KW Sensory transduction; Transferrase; Kinase; Phosphorylation;
 KW Transmembrane; Inner membrane; Complete proteome.
 FT DOMAIN 1 21 CYTOPLASMIC (POTENTIAL).
 FT STRAINMEM 22 182 POTENTIAL.
 FT DOMAIN 43 482 PERPLASMIC (POTENTIAL).
 FT STRAINMEM 183 203 POTENTIAL.
 FT DOMAIN 204 552 CYTOPLASMIC (POTENTIAL).
 FT STRAINMEM 222 292 PAS.
 FT DOMAIN 344 541 HISTIDINE KINASE.
 FT MOD_RES 347 347 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

FT CONFLICT 518 552 GGVTLIEDNPGTLESITYPKKPNDSINPIDR -> VV
 FT SEQUENCE 552 AA: 61684 MW: 383F0E081ADB6E77 CRC64;
 SQ SEQUENCE 552 AA: 61684 MW: 383F0E081ADB6E77 CRC64;
 Query Match
 Best Local Similarity 21.5%; Score 128.5; DB 1; Length 552;
 Matches 90; Conservative 81; Mismatches 150; Indels 97; Gaps 20;
 QY 18 VPELFLEFSKVSNTVLSKELTLESISNPLIMAVTWNVNFPA-----EPLVF 68
 DB 162 VVISGLVSKIDS--WAEELPMAVEYVLLGLML-LSWFLAAHRRQMGMEPK-- 215
 QY 63 IALSTYLNKNSLSLNIFFGLPVAASDLPFRRAIFFLDGTGIVGSSITTYAIEFA 128
 DB 216 -GLARVVRQGEALFSSVEYEGLLAV-----DPHGYYIAINRNKKNLGLS 258
 QY 129 GIALSYLPLSVNNDIGLKSILTKMKYKRLIMNTMLLYLLIOVLYIESVNIPL 188
 DB 259 SPGRKWLGRPI--VEVVRPADFTFEOIDEKR-----QDVANFNGLSV 299
 QY 189 LKFRFVVIYVLLIFLLISFLSQYTKOKVNEIMAKREAIRNITQYSSQIESLYKDIR 248
 DB 300 IANREAIRSGDGLGAL-ISFRSK-----DEISLNMQLQIKY--VESL-----R 343
 QY 249 SRRHVYMLNISLRIGIKNKDLASIEKTYHOILEKTHQLODTRYNIGHLANIQND-AYK 307
 DB 344 YLRHEHLMWSTLNLQMKR-----YDRVLA MVGEGESAOQDLDSIREAFAPROVA 396
 QY 308 GILSAKILEAOKKIAVAVESSKI-OLPE-MELLDEYTLISLDCNAIAFAES--LNP 363
 DB 397 GLLPKQVQARRELKMIIVGSQLSQLPGLDSTEFALVGNLDAFASLRSDGK 456
 QY 364 ETQLAFEFKKNISYFI-----IQNSTKQIDVSKIKRENTSTKG---SKRGIGL 410
 DB 457 IYEL-FLSDEGDVYIEVADGCGVPESLRD-----KIEGQVSTRADEPGEHIGL 507
 RESULT 6
 CHGV_RHIME STANDARD; PRT; 577 AA.
 ID CHGV_RHIME STANDARD; PRT; 577 AA.
 AC P72292; 031069;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 OS Sensor protein chvg (EC 2.7.3.-) (Histidine kinase sensory protein
 exos).
 DE CHGV OR EXOS OR R00043 OR SMC04446.
 OS Rhizobium melioli (Sinorhizobium melioli).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 OC NCBI_TaxID=382;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-1021;
 RX MEDLINE-98083050; PubMed-9422587;
 RA Cheng H.-P., Walker G.C.,
 RT "Succinoglycan production by *Rhizobium melioli* is regulated through
 the ExoS-chv1 two-component regulatory system.";
 RL J. Bacteriol. 180:20-26(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-1021;
 RX MEDLINE-21396507; PubMed-11481430;
 RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Anpe F., Batut J.,
 RA Bolstad F., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
 RA Godtke T., Goffeau A., Kahn D., Kiss E., Lelaur V., Masny D.,
 RA Pohl T., Portetelle D., Puehler A., Punelle B., Ramsperger U.,
 RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.,
 RT "Analysis of the chromosome sequence of the legume symbiont
Sinorhizobium melioli strain 1021.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
 RN [3]

```

Db      446 KDNPRASFVYSGEALRGQIIINLEMA-----RSFVDE-----ONCRIVYLRLSL 493
OY      375 -SIVEIITNSTRKQIDVKIFKEKNYSRK-----GNSRNGGLAKVHILHEHPKTSLOT 427
       ||:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
Db      494 RCIYVEEDNPGIQADIDIRIFERYTTDRPEGEDFGONSGLGISISQRIEAHGSI-LRA 552
OY      428 SN 429
       |
Db      553 EN 554

RESULT
7
DCTS_BACSU STANDARD: PRT: 535 AA.
AC p96601;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Problem C4-dicarboxylate sensor kinase (Ec 2.7.3.-).
GN DCTS.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA MEDLINE=98000887; PubMed=9341680;
RA Beutin C., Ayora S., Exley R., Hirschbein L., Ogasawara N.,
RA Kasahara Y., Alonso J.C., Le Hegarat F.,
RT "Characterization of an lrp-like (ltpc) gene from Bacillus subtilis.";
RM Mol. Gen. Genet. 256:63-71(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP FUNCTION, AND GENE NAME.
RC STRAIN=168;
RX MEDLINE=20170658; PubMed=10708364;
RA Aasi K., Balk S.-H., Kasahara Y., Moriya S., Ogasawara N.;
RT "Regulation of the transport system for C4-dicarboxylic acids in
RL Bacillus subtilis."
RM Microbiology 146:263-271(2000).
CC -1- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM DCTS/DCTR.
CC PROBABLY ACTIVATES DCTR BY PHOSPHORYLATION (BY SIMILARITY).
CC ESSENTIAL FOR EXPRESSION OF DCPT.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.
CC -----
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
Cc      EMBL; AB001488; BAIA19282.1; -
Cc      EMBL; Z99106; CAB12252.1; -.
Cc      Subtilist; BG12073; dcts.
Dr      InterPro; IPR003594; HARPAse_C.
Dr      InterPro; IPR004359; HIS_KIN_Sig.
Dr      InterPro; IPR000014; PAS.
Dr      Pfam; PF02518; HARPAse_C_1.
Dr      Pfam; PF00512; signal; 1.
Dr      SMART; SM00387; HARPAse_C_1.
Dr      SMART; SM00091; PAS; 1.
Dr      PROSITE; PS50109; HIS_KIN; 1.
Dr      PROSITE; PS50112; PAS; FALSE_NEG.
Tr      Sensory transduction; Transferrase; Kinase; Transmembrane;

```

KW Phosphorylation; Complete proteome.
 FT DOMAIN 1 11 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 12 32 POTENTIAL.
 FT DOMAIN 33 172 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 173 193 POTENTIAL.
 FT DOMAIN 194 535 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 213 276 PAS.
 FT DOMAIN 333 528 HISTIDINE KINASE.
 FT MOD_RES 336 336 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 SQ SEQUENCE 535 AA; 59942 MW; 614BD8B8F10CFC CRC64;

Query Match 5.8%; Score 126.5; DB 1; Length 535;
 Best Local Similarity 18.7%; Pred. No. 1.1;
 Matches 84; Conservative 91; Mismatches 137; Indels 137; Gaps 24;

QY 63 AEPLVIALSLYLNQNSLSLNFYGLPLVASSDLFRRAIFFLDGTQGVMSSTITT 122
 DB 121 AEHIFSE---AKGEIGTAVRAFY---PVKDQDLNQG---VLVQKTL--- 160
 QY 123 YMEFAGIA--LSYL-----FLSVFNVDIGRLKSLTKMKYKRLIPMNITMLLYLLIQ 175
 DB 161 ----PGIADILLHLKRDIAFIVLTGLGSLAGSFLARHIKKO----- 199
 QY 176 VLVIESYVIV-----PLAKFKREVIVYLLFLFLISFLSQYTKO--KYONEIMAO--- 225
 DB 200 -MEQLPHEIVKMEBERTATFSMNEGVIAIDNRLVITTEFNKAKQFEVQGLIGKVIW 238
 QY 226 --KEAOI-----RNITQYSOQIE-----SLYKD----- 246
 DB 259 EVLKDSRLPEIVERNKAVNEIRVSKVIMSSRIPIYMKKVIQAVAFEDPTEAKMA 318
 QY 247 -----RSF-----RHLYNLITS-----RLGIENKDLASIEKIYHQLKRGHLOD 290
 DB 319 EELTGKRVNEALRVONHEHMKHTIAGLIOLGKSEKAL-----QLAFDAST 366
 QY 291 TRYNIHL--ANIQNDVAKGILSAKILEAONKKIAVANVESSKI-QDPE-MELDFTTIL 346
 DB 367 EEDNVEFEFLHSIQNDAAAGLLSLKIRGRELGIIVHIDENSSIQDFEHVDQHDIVLL 426
 QY 347 SIICDVAIAFAFESLNPE---IQLAFPKKNGSIYFIIO--NSTKEQIDVSKIFEENYS-T 401
 DB 427 GNLEENAF-GSEFTYQSEKRIIDISIEQTDLLALILEDNCGIEPHMPALYDKGFTVN 465
 QY 402 KGSNKGIGLAKVNHILEHYPKTSIQTSNH 430
 DB 486 KTGCTGYGLYVKQIIDKSGST-IEVDSH 513

RESULT 8
 LCN3_LACLA STANDARD; PRT; 691 AA.
 ID LCN3_LACLA
 AC P37608;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Lactacin 481/lactococcin transport ATP-binding protein lcnbr3.
 GN LCNBR3.
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
 OC Bacteria; Firmicutes; Bacillales; Clostridium group; Streptococcaceae;
 OC Lactococcus.
 OX NCBI_TaxID=1360;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-ADRIA 85LO30;
 RX MEDLINE-94288641; PubMed-8017945;
 RA Rince A., Dufour A., le Pogam S., Thuaud D., Bourgeois C.M.,
 Rancec J.P.;
 RT "Cloning, expression, and nucleotide sequence of genes involved in
 production of lactococcin DR, a bacteriocin from Lactococcus lactis
 subsp. lactis.";
 RT Appl. Environ. Microbiol. 60:1652-1657(1994).
 CC -1- FUNCTION: PROBABLY IMPLICATED IN THE EXPORT PROCESS OF THE

CC LANTIBIOTIC LACTICIN 481/LACTOCOCCIN DR.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. HLYB SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb-sib.ch).

DR EMBL; U91581; AAC72259.1; -
 DR InterPro: IPR003593; AAA.
 DR InterPro: IPR001140; ABC transporter_tmem.
 DR InterPro: IPR003439; ABC transporter.
 DR InterPro: IPR001687; ATP_GTP_A.
 DR Pfam; PF00664; ABC_membrane; 1.
 DR Pfam; PF00005; ABC_tran; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
 KW Transport; Protein transport; Bacteriocin transport; Transmembrane;
 KM ATP-binding.
 FT TRANSMEM 157 179 POTENTIAL.
 FT TRANSMEM 189 208 POTENTIAL.
 FT TRANSMEM 262 284 POTENTIAL.
 FT TRANSMEM 289 311 POTENTIAL.
 FT TRANSMEM 381 403 POTENTIAL.
 FT NP_BIND 497 504 ATP (POTENTIAL).
 SQ SEQUENCE 691 AA; 79834 MW; 4276DD778AEC0B47 CRC64;

Query Match 5.8%; Score 126.5; DB 1; Length 691;
 Best Local Similarity 19.5%; Pred. No. 1.5; Indels 149; Gaps 25;
 Matches 103; Conservative 96; Mismatches 179;

QY 14 TYLVFLFLFL-----FSKVSNTLSKKELFLFSISNPLIMAVYV----- 55
 DB 158 TFL-YIFSLFISQIALVAFSLIIDLINKSHDIYS--PIMMSHVLPTSLMKLGA 213
 QY 56 --NVLFPAELVFLALSLYLNQNSLSLNFYGLPLVASSDLFRRAIFFLDGTQGI 113
 DB 214 QKNTMLY-----ESKISROIENG-----IFSRLYF-RNNSVGT 248
 QY 114 VNGSSITTYMIEFAGIALSYLFSVFNVDIGRLKSLTKMKYKRLIPMNITMLLYLL 173
 DB 249 IEKINLRGTGRD--GILTK-IFPSLNF-----FVFEIVLYIGTISFLTLFLVI 297
 QY 174 IOYLVIESYVIVPLTKRKRFVIVYLLFLFLISFLSQYTKOYONEIMAOKEAOIRNI 233
 DB 298 KMLTYMIFSFSL--SIKROANIQTQITIDFTSVQEDLNQIDQIKAOANEKECVKRW 354
 QY 234 TOYSOQIESLYKDIRSF-----RHLYNLITSRLGI--ENKDLASIEKIYHQLIE 282
 DB 355 TKKSAQITTFYVYKILINDIGITSANQGFNYICVILAMIFGIYVNOGLVSIPL--IIF 411
 QY 283 KTG-----HOLDPEYVNIHGLA-----NIQNDVAKGILSAKI-- 314
 DB 412 QSGISLPSAVNQIDVAFELSRSLSTYGNKISDLLENPORDINIEKHSNMAIILNDISY 471
 QY 315 -LEAON---KIAVANVESSKIQLPEMELDFTTILSIIC-----DNAIEAFASINPEI 365
 DB 472 SYELNNYIFNNINFSIKKGEKIAIVGKSGSKSTFLNILLGLISREGVYGYEENRQIT 531
 QY 366 QLAFFKKN--GSIVFIIONSTKKEQIDV-----SKIKENYSTK 402
 DB 532 GYVSQNMMLRKSGSLIENIVSNNSNEELDQKINDVLDVAMMELVYSLPKIFISQLE-N 590
 QY 403 GSNKGIG-----IAK--VNH---ILEHYPKTSIQTSNHHHFLKQL 438
 DB 591 GKNLSGQIQIRLLAKSLNNKKEIFWDEPSSLDNQRNIHYKNYL 637

```

RESULT 9
MPE2_MSEPV STANDARD: PRT: 717 AA.
ID MPE2_MSEPV
AC Q9YK06;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nucleoside triphosphatase II (EC 3.6.1.15) (NTPase II) (Nucleoside
triphosphate phosphohydrolase II) (NPH II).
GN MSY086;
OS Melanoplus sanguinipes entomopoxvirus (MSEPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;
OC Entomopoxvirus B.
OX NCBI_TaxID=83191;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tucon;
RX MEDLINE=99102612; PubMed=9847359;
RA Alfonso C.L., Tulman E.R., Lu Z., Oma E., Kutish G.F., Rock D.L.;
RT "The genome of Melanoplus sanguinipes entomopoxvirus.";
RL J. Virol. 73:533-552(1999).
CC -1- FUNCTION: ESSENTIAL FOR VIRAL REPLICATION. PLAYS AN IMPORTANT ROLE
CC DURING TRANSCRIPTION OF EARLY MRNAs, PRESUMABLY BY PREVENTING R-
CC LOOP FORMATION BEHIND THE ELONGATING RNA POLYMERASE. ACTS AS NTP-
CC DEPENDENT HELICASE THAT CATALYZES UNIDIRECTIONAL UNWINDING OF
CC 3'TAILED DUPLEX MRNAs. MIGHT ALSO PLAY A ROLE IN THE EXPORT OF
CC NEWLY SYNTHESIZED MRNA CHAINS OUT OF THE CORE INTO THE CYTOSOL.
CC REQUIRED FOR PROPAGATION OF VIRAL PARTICLES (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: NTP + H(2)O = NDP + phosphate.
CC -1- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY. DEAH
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: AF063866; AAC97810.1; -.
CC DR InterPro: IPR002410; DEAD.
CC DR InterPro: IPR002464; DEAH_ATP_helicase.
CC DR InterPro: IPR001650; Helicase_C.
CC DR Pfam: PF00271; Helicase_C.1.
CC DR SMART: SM00490; HelicC; 1.
CC DR PROSITE: PS00690; DEAH_ATP_HELICASE; FALSE_NEG.
CC DR ATP-binding; Helicase; Hydrolase; Transcription.
CC FT DOMAIN 328 533
CC FT NP_BIND 206 213 ATP (POTENTIAL).
CC FT SITE 331 334 DEAD BOX.
CC FT SEQUENCE 717 AA; 84725 MW; E8FB04661F1EAB63 CRC64;

```

```

Query Match 5.88; Score 125.5; DB 1; Length 717;
Best Local Similarity 18.33; Pred. No. 1.8;
Matches 107; Conservative 97; Mismatches 172; Indels 209; Gaps 24;

```

```

QY 15 YLTVLFLFLFKSVNNTLSKKEFLFSLNFMIAVTWNVNLPAPAEPLFALSIY 74
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 26 YVYMAVLLF---PNNATFNSYITKEVEFYPMQFAL- - - - -PYVKLTWHNINIC 75
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 75 LNR-----QNSLSINIFYGLPLVASDLPFRRAIF-----FILDG----- 109
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 76 LNNRFYLSNEFKNNISINTVYNL-----LYNNELKFEDDNIINGKMLKISYASY 128
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 110 ---TGGIYWGSSITTTTMI-----EPAGTALSYPLEPVNDIGLAKSLKMKYKR 139
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 129 VTIISQITINISLNKRYQIYGIIESANLYGILSSKONKRYFDKNL-----FSFKSELRST 184
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 160 LIPNMITMLLYL-----LIQVLY-----VIESY----- 183
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 185 MLDVQLKIFELFISKKNCISGCTGCTGTYIVPLKFWMFNLLPFGYEFWMVSNKNKIND 244
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

QY 184 -----NVITLTKPRK-----VVIY----- 199
DB 245 FIFKPNFEKNTILSLPRKALIRMGINTKSLGFDNISGSPITLLKYKKKEKEYNN 304
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 200 -LILFLILISFLS---OYTKOKVONEI-----MAQEAQIRNITQYQ 238
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 305 PLVPEVLSVNRITINNIKHSNSYITDEIHEHDFGDIAIARFKKKRYNIRNIVLSA 364
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 239 QIESLYVDIRSEFHDYINILTSRLGI-----ENKDLASIEKIYHQLERTGQL 288
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 365 TIES---DIDIRIKRYENIYIIPGVSLEPVEKECEKQVSIKKNMPVSGSVIIF 421
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 289 QDTRNIGHLANIOMDAVKGILSAKILEAO--NKIAVANVESSKQTLPDEMELDTITL 346
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 422 YETIKI---NEKKEILESILIDIKYKITYIKTSKITTINAINAKIQNDKKH-----IH 471
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 347 SLIDGNIEAFESLNEPEIOL---AFKK--NGSIVFIIONSTKEKOLDYKIFRENS 400
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 472 IILSTINIESSITTTNNTVIDNGMKYKKKFLGSGTMYITESMYIORKRGVRISKGY- 530
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 401 TKGSRGIGLAKVNHLEHYPKTSLSQTS---NHHHLFROLIIR 441
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 531 -----IRYSKDLQTTFKHINYQYLMETIIVEK 559
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 10
V120_HSV7J STANDARD: PRT: 938 AA.
ID V120_HSV7J
AC P52438;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Capsid assembly protein U30.
GN U30.
OS Human herpesvirus (type 7 / strain J1) (HHV7).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=57278;
RN [1]
RP SEQUENCE FROM N.A.
RA Nicholas J.;
RT Submitted (JUN-1996) to the EMBL/GenBank/DDI databases.
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL37,
CC HSV-1 23, EBV BOLF1, VZV 21, HSV-1 63, AND HCMV UL47.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: U43400; AAC54692.1; -.
CC DR Capsid assembly.
CC FT SEQUENCE 938 AA; 110170 MW; F4E39A2BF0D32BC9 CRC64;

```

```

Query Match 5.88; Score 125.5; DB 1; Length 938;
Best Local Similarity 20.38; Pred. No. 2.5;
Matches 92; Conservative 70; Mismatches 112; Indels 179; Gaps 22;

```

```

QY 2 NEALMIL-----SNGLTLYTVLFLFLFKSVNNTLSKKEFLFSLI----- 43
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 577 NRRLLILFLACKLMPNSVILSH--YLLLLHART-----LDIFVVDGHSIIHAITQKIF 630
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 44 -----SNFLIMAVTMVNVNLFYPAEPLFALSITYNR--QNSLSINIFY 87
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 631 DNINSLOTPIPTNPLVSLILF-----AYVHMQTYVNPVIOKTISENI-- 676
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 88 GLPLVASDLPFRRAIFILDGTGGIYWGSSITTYIIEFAGTALSYP-----LS 138
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 677 -----ALRKEYIDFTK--KCSSTLATTCYLNLENFAVMYFGKKNKGTSLS 721
QY 139 VENVDIGRL-----KESLTKMKYKRRLLPNMTLLYLLIQVLYIEBYNIPFL-K 190
      :|: | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 722 AFRICCSKILIESKLEPKRLOELVSKTLF--TEML-----QNNVKKITK 764
      :|: | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 191 FRKVVIVYLLLELLISLSQYTKQVONEIMAEQIRINITYSOOISLKYDIRSF 250
      :|: | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 765 FKDLV-----SNQTLQPII-----VERISSHANTYQDV--- 795
      :|: | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 251 RHDYUMLTSLMGLKENDKLASIEKIYHQLLEKTHQLODPTVNGIHLANIONAVGI- 309
      :|: | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 796 -----LMSIDECF-----SNMQLQSKRNIVYIVIDLN--TNIF 829
      :|: | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 310 ---LSAKILLEAONKRIAVNEVSRIOLPEMELLDITLLSICDNAIEAASFLEIQ 366
      :|: | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 830 NELSALQLEIA--KRLVKKQDPTYNQULWQD---DEPVILKSLNNLPEKOKPTIINTER 883
      :|: | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 367 LAFFKKNSIVFIITONSREKQIDSVKIFPENY 399
      :|: | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 884 -----FMELGSDITKQIQLPFLDVEPERY 905
      :|: | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

Db      7  IFEYIASSIFUYLAKNOVDNI-----TALVNDIELNSVDNEILVELFKSSKKFIIPLK 61
QY      211  SÖTTRKQVONELMAO-----KKAQNRITQYSOQIESJYDINFSRH---DYLNTILTSRL 263
      62  SDPLKATVELEKILVDLSLLDQNSKNRINITYKEQIDIVYKMLKKNHSYDHFQKQILR- 120
QY      264  GIENKCLAS--TEKTYHOILEKTHQODTRVYNGHLANIQNDAYKGLISAKILAEONKR 321
      121  NIKNPYVONPFIKKEITELMKKT---IQD--VELHKRINISQEVNTIFK--KLKDNKEF 174
Db      322  IAVNVESSKIQDPEMLLDPITLITSLCONALEA-----FESLPEI 365
QY      175  KRIHL--SYLLPBLKODS-----DNVAKNRKTAENIYKLLKGIDYEFKLLIEC 222
Db      366  ÖLAFEFKNGSIVELIÖNSTKEQIDVSKIFE--NYSTQ-----SNRGIGLAKVNYH 416
      223  E-----KNKS-TPYIAKKMFWPRLDIONSEFFKTLTFKKGQILGPVBDGKGYILIKVNDI 276
QY      417  LEHYPTSLÖTSNH--HHLPKQOLLII 440
      277  --HHKENIYTEFYMQHCLIKRPVIL 300
Db

```

SURRA	BUCAI
ID	SURRA_BUCAI
AC	p57240;
DT	16-OCT-2001 (Rel. 40; Created)
DY	16-OCT-2001 (Rel. 40; Last sequence update)
DT	16-OCT-2001 (Rel. 40; Last annotation update)
DE	Survival protein surra homolog precursor.
GN	SURA OR BU140.
OS	Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum symbiotic bacterium).
OC	Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX	NBIL_taxonomy=118099;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	SIRRAIN-TOKYO 1998;
RA	MEDLINE-20445173; PubMed-10993077;
RT	Shigenobu S., Watanabe H., Hatton M., Sakaki Y., Ishikawa H.; "genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS."
RL	Nature 407:81-86(2000).
CC	- SIMILARITY: BELONGS TO THE PPIC/PARYULIN FAMILY OF ROTAMASES.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed, usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/atlib-sib.ch).
CC	or send an email to license@isb-sib.ch).
DR	EMBL; AP001118; BABI2858.1; -
DR	InterPro; IPR000297; Rotamase:
DR	Pfam; PF00639; Rotamase; 1
DR	PROSITE; PS01096; PPIC_PPIMSE_1; FALSE_NEG.
KW	PROSITE; PS50198; PPIC_PPIMSE_2; 2.
KW	Isoomerase; Rotamase; Repeat; Signal; Complete proteome.
FT	SIGNAL
FT	CHAIN
FT	DOMAIN
FT	DOMAIN
SO	SEQUENCE

```

RESULT 12
RPOB_PLAFA
ID RPOB_PLAFA STANDARD; PRI: 1024 AA.
AC p21421;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE DNA-directed RNA polymerase beta chain (EC 2.7.7.6).
GN RPOB.
OS Plasmodium falciparum.
OC Eukaryotes; Alveolates; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5835;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BM(C10);
RX MEDLINE=95107345; PubMed=7808472;
RA Gardner M.J., Goldman N., Barnett P., Moore P.W., Rangachari K.,
RA Strath M., Whyte A., Williamson D.H., Wilson R.J.M.;
RT "Phylogenetic analysis of the rpoB gene from the plasmid-like DNA of
RL Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 66:221-231(1994).
RN [2]
RP SEQUENCE OF 328-1024 FROM N.A.
RX MEDLINE=91187055; PubMed=2011147;
RA Gardner M.J., Williamson D.H., Wilson R.J.M.;
RT "A circular DNA in malaria parasites encodes an RNA polymerase like
RT that of prokaryotes and chloroplasts.";
RL Mol. Biochem. Parasitol. 44:115-124(1991).
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate + N diphosphate +
CC [RNA](N).
CC -1- MISCELLANEOUS: THIS RNA POLYMERASE IS ENCODED ON A CIRCULAR DNA.
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/)

```

```
Query March      5-5% Score 120.5; DB 1; Length 430;
Match Local Similarity 23.0%; Pred. No. 2;
Matches 75; Conservative 55; Mismatches 111; Indels 85; Gaps 19.

Oy 168 LLYVLLIQGVYVESNVNPTLKRFRVVIV-----YLTLFL-----ILSLF 210
       :|::: ||: |::: | |
```

DR EMBL: X75544: CAA53232.1: -.
DR EMBL: X52177: CAA36427.1: -.
DR EMBL: X95275: CAA64572.1: -.
DR PIR: S10438: RNZOBP.
DR InterPro: IPR001572: RNA_pol_B.
DR Pfam: PF00562: RNA_pol_B.1.

FT TRANSMEM 129 151 POTENTIAL.
 FT TRANSMEM 158 175 POTENTIAL.
 FT TRANSMEM 179 196 POTENTIAL.
 FT TRANSMEM 203 225 POTENTIAL.
 FT TRANSMEM 304 326 POTENTIAL.
 FT TRANSMEM 347 369 POTENTIAL.
 SQ SEQUENCE 388 AA; 45529 MW; F06E7BEEB1B9941 CRC64;

Query Match 5.4%; Score 117.5; DB 1; Length 388;
 Best Local Similarity 22.4%; Pred. No. 2.6;
 Matches 80; Conservative 55; Mismatches 127; Indels 95; Gaps 15;

QY 2 NEALM-----ILNSGLITLYLVLFLFESKVSNTLSKELT-----LFSISNPLMI 50
 DB 7 NEULLALVYSAFLSISLEGVYIITLLVLYKIKERKIKOSLPGLILSL----- 60
 QY 51 AATWVNVNLEFP-----AEPYLTALSIYLRON-----SLSLNTFYGLLPVAS 94
 DB 61 --TVLSTAFYPRKFLKIEEGLFQFIYFLNLKKEVGFSGKIFPKLLIGLILPVVF 118
 QY 95 SDFERRAIFPIDGT--OGIVGSSITTYMI-----EPAGIALSYLFLSVFVNDIGRL 147
 DB 119 YKFKYKGEPRPIMGTFEYGFPALEFISITTLFFKERRFTIPLFLFLAV----- 170
 QY 148 KSLTFKAKYKRLIPNITMLLYLLIOVLYIESYNIAPLTKRREVIIVY-LLEFLIL 206
 DB 171 -----IFLSARRSMIAFFVIFYLFLVLF-----KSKKIKLAFWSVNFLLI 213
 QY 207 ISFLSYOT-----KOKVONEIMAKOEQIRNITQYSOQIESLKDRSPRHVDNLTL- 258
 DB 214 LSLFIGGVYLSQKHFKFLNDLILGK-----BLNQNLNLSISSGGLNLL 260
 QY 259 ----TSLRLGLENKDLASIEKIYHQL--EKGHQLQDTRYNIGHLANIQDAVKCIL 310
 DB 261 EGISIIKEDIENKRFNL-LIGHGVAGEMPHRGYMOHREYSIFVSEFLERGITL 316
 RESULT 15
 YC26_PORPU STANDARD; PRT; 656 AA.
 AC PS1392;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Hypothetical sensor-like histidine kinase ycf26 (EC 2.7.3.-).
 GN ycf26.
 OS Porphyra purpurea.
 OC Chloroplast.
 OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
 OX NCBI_TaxID=2787;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AVONPORT;
 RA Reith M.E., Munnholland J.;
 RT *Complete nucleotide sequence of the Porphyra purpurea chloroplast genome.*
 RL Plant Mol. Biol. Rep. 13:333-335(1995).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN; CHLOROPLAST (PROBABLE).
 CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL; U38804; AAC08278.1; -
 CC InterPro; IPR004358; BCTRLSENSOR.

DR InterPro; IPR003660; HAM.
 DR InterPro; IPR003594; HATPase_C.
 DR InterPro; IPR004359; HIS_KIN_sig.
 DR InterPro; IPR003661; His_kinA.
 DR InterPro; IPR000014; PAS.
 DR Pfam; PF00672; HAM; 1.
 DR Pfam; PF02518; HATPase_C; 1.
 DR Pfam; PF00989; PAS; 1.
 DR Pfam; PF00512; signal; 1.
 DR PRINTS; PR00344; BCTRLSENSOR.
 DR SMART; SM00304; HAM; 1.
 DR SMART; SM00387; HATPase_C; 1.
 DR SMART; SM00388; HISKA; 1.
 DR SMART; SM00091; PAS; 1.
 DR PROSITE; PS50109; HIS_KIN; 1.
 DR PROSITE; PS50112; PAS; 1.
 KW Hypothetical protein; Sensory transduction; Transferase; Kinase;
 KW Transmembrane; Phosphorylation; Chloroplast.
 FT TRANSMEM 7 27 POTENTIAL.
 FT TRANSMEM 40 60 POTENTIAL.
 FT TRANSMEM 210 230 POTENTIAL.
 FT DOMAIN 289 359 PAS.
 FT MOD_RES 424 654 HISTIDINE KINASE.
 FT MOD_RES 427 427 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 SQ SEQUENCE 656 AA; 75362 MW; 2B3DDE7C91AE680 CRC64;

Query Match 5.3%; Score 116; DB 1; Length 656;
 Best Local Similarity 20.5%; Pred. No. 5.9;
 Matches 68; Conservative 68; Mismatches 110; Indels 86; Gaps 15;

QY 161 IPMNITMLLYLLIOVL-----YVESYVIFTLK--FRK-FVYVYVIL--FLIL- 206
 DB 340 LPEDINQQLFPLINDIRKNFLBOSICETQICIKLNKTKKTFVLTVDHKYSILK 399
 QY 207 -ISFLSYTKKQVQ-NEIMAKOEQIRNITQYSOQIESLVKDIRSF-----RHAYNLIL 258
 DB 400 GIAMTQDRTQVEVLEIKNQ-----FISVSHELRTPLEINRSFETLYEYHSDLDS 453
 QY 259 TSLRLGLENKDLASIEKIYHQLLEKTHQQLQDTRYNIGHLANIQND-----A 305
 DB 454 QKLEPLAANKETGRILVNDVLD-----LSRLSDQETTLQPTDLVSA 498
 QY 306 VGLISAKILEAKNKIIVANVEVSSKIQLPPEMLDFTLLSLCDNAIPA----- 356
 DB 499 VQGITRTYQLSAKDRIDLDIDIEONLQ--CVLGNVLLQILANLVVNSLFTAPNGI 555
 QY 357 -----AFESLNPETQLAFPKKNGSIYFIIONSTRKEROIDVSKIFK-----ENYSTKGSN 405
 DB 556 ILKRAYVDLKTETEVQHFNSQKRYVEICDNGIGISRKQDERIFAFRLIENYVHTLGS 615
 QY 406 RGIGLAKVNHILIEHYPKTSLOTSNHHHLKOL 437
 DB 616 TGIGLSIVAKNIQKH-----NSEIHLXSEL 640

Search completed: July 30, 2002, 15:20:54
 Job time: 365 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:20:30 ; Search time 45.14 Seconds

(without alignments)
1690.092 Million cell updates/sec

Title: US-09-833-017-4
Perfect score: 2173
Sequence: 1 MNEALMILSNGLITLVLF.....KTSLSQTSNHHFLKQLLIK 441

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Listing first 45 summaries

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protist:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaea:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	773	35.6	446	2	09F2F2 streptococ
2	755	34.7	446	2	09F2F5 streptococ
3	749	34.5	446	16	09S115 streptococ
4	545.5	25.1	426	2	032832 streptococ
5	539.5	24.8	448	16	09A1J7 streptococ
6	526	24.2	451	16	099YPA streptococ
7	517	23.8	427	16	09A1J6 streptococ
8	455	20.9	447	16	097M11 clostridium
9	455	20.9	447	2	P95767 streptococ
10	450	20.7	453	2	P72446 streptococ
11	449.5	20.7	453	2	033646 streptococ
12	443.5	20.4	448	2	033597 streptococ
13	439.5	20.2	441	2	09X995 streptococ
14	438.5	20.2	441	2	09X990 streptococ
15	437	20.1	452	2	P72444 streptococ
16	436.5	20.1	440	2	033616 streptococ

17	436.5	20.1	441	2	033652 streptococ
18	435.5	20.0	441	2	033667 streptococ
19	434.5	20.0	441	2	033669 streptococ
20	432.5	19.9	441	2	09X9B8 streptococ
21	427.5	19.7	440	2	033671 streptococ
22	427.5	19.7	441	16	09S1K4 streptococ
23	426.5	19.6	441	2	09X9B6 streptococ
24	426	19.6	441	2	033673 streptococ
25	424.5	19.5	441	2	09X9B4 streptococ
26	423.5	19.5	441	2	09X9B2 streptococ
27	420.5	19.4	441	2	09X9B2 streptococ
28	416	19.1	431	16	Q92FRO listeria in
29	414.5	19.1	441	2	P72566 streptococ
30	411.5	18.9	429	2	09S3T7 streptococ
31	394	18.1	425	2	033590 streptococ
32	389	17.9	437	2	033587 streptococ
33	381	17.5	430	2	052182 streptococ
34	379	17.4	423	2	033644 streptococ
35	378	17.4	411	2	09F622 streptococ
36	378	17.4	429	2	068159 streptococ
37	377	17.3	430	2	007911 streptococ
38	375	17.3	425	2	09F1U4 streptococ
39	375	17.3	425	2	09F1U2 streptococ
40	375	17.3	430	2	052180 streptococ
41	374	17.2	430	2	052181 streptococ
42	372	17.1	371	16	Q93SL2 streptococ
43	366.5	16.9	427	2	09X473 streptococ
44	362	16.7	428	2	048866 lactobacill
45	362	16.7	430	2	048874 lactobacill

ALIGNMENTS

RESULT 1
ID Q9F2F2 PRELIMINARY; PRT; 446 AA.
AC Q9F2F2;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HISTIDINE KINASE.
GN HK.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sp670;
RX MEDLINE=20487097; PubMed=11034284;
RA Reichmann P.; Hakenbeck R.;
RT "Allelic variation in a peptide-inducible two-component system of Streptococcus pneumoniae."
RL FEMS Microbiol. Lett. 190:231-236(2000).
DR EMBL: AJ278302; CAC10175.1; -
DR InterPro: IPR001064; Crystal11n.
DR InterPro: IPR003594; HATPase_C.
DR SMART: SM00387; HATPase_c; 1.
DR PROSITE: PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
KW Kinase.
SQ SEQUENCE 446 AA; 52057 MW; 98065278BB9C2C1E CRC64;

Query Match 35.6%; Score 773; DB 2; Length 446;
Best Local Similarity 38.8%; Prod. No. 5.4e-37;
Matches 174; Conservative 91; Mismatches 168; Indels 16; Gaps 6;

QY 1 MNEALMILSNGLITLVLFLLFSKVSNTSKKELFLFSNPLIMAVT-----M 54
DB 1 MNVAAIL-----LYTLVTHGKLIVIFKVDGISLTFER--IFKAFLEKILLAVFGMLGYM 54
QY 55 V-NVALFPAPPLVIFALISVILNQNLSLINFYGLLPVASSDLFPRAILFIID--GTQ 111

```

Db 55 VGNVYLSFMPRLVIGIGISLLELRLKLLRGLFPLMLVNLFRGVSYPFLPLGGG 114
Oy 112 GIVMGSSITTYTMEFAGIALSYLSEVNDIGRLKDSLTKMKVKRLIPMNTMLLY 171
Db 115 QYVDYSFIMCIIIF--NEFISLAFKMLDYDFSLRKGLDDEPKSLQJIMWAGAY 173
Oy 172 LLIQVLYIESYVNPITLKFKEFYVYIYLLFLLISFSLQYTKQVQNMOKRQAOJR 231
Db 174 LVIQMLSFEEYQGIQSTTVHLLVYLLFPMGIKLLTYLKDKLHELNQDRLR 233
Oy 232 NITQYQOIESLYKDIRSRHDYINILTSRLGIENKDLASIEKIYHQILEKTHQLODT 291
Db 234 EMERSRHIEELKYEVSRFDYTNLLTSLRLGIEEDMQIKREIYDSVLKSSSEKLDN 293
Oy 292 RYNIHLANIONDAVKGLSLAKIIEAQNKKIYAVNEVSSKIQLEPEMLDFTILSLCD 351
Db 294 KYDGLRVNDRALKSLACKFKAKDKNIVFVEPEIYQEGVSLDELTVVSLD 353
Oy 352 NAIEAFESLNPETQLAFKKNGSIVFIQNSTKEKOIDVSKIFKENYSTKSGNRIGLA 411
Db 354 NAIEASVACOPHVAIFKNGAOTETFIENSIRKEGIDISEIFSGASSKGEERGVLTY 413
Oy 412 KVNHLHYPKTSLOTSNHHLLFKOLLII 440
Db 414 TVMKIVESHPTSLNTGQDHVFRQVLTV 442

RESULT 2
Oy 255 PRELIMINARY; PRT; 446 AA.
ID 09F2F5
AC 09F2F5
DT 01-MAR-2001 (TReMBLrel, 16, Created)
DT 01-MAR-2001 (TReMBLrel, 16, Last sequence update)
DT 01-DEC-2001 (TReMBLrel, 19, Last annotation update)
DE HISTIDINE KINASE.
GN 1H.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R6;
RA MEDLINE=20487097; PubMed=11034284;
RA Reichmann P., Hakenbeck R.;
RA "Allelic variation in a peptidase-inducible two-component system of
RT Streptococcus pneumoniae.";
RL FEMS Microbiol. Lett. 190:231-236(2000).
DR InterPro; IPR001064; Crystalin.
DR SMART; SM00387; HATPase_c.1.
DR PROSITE; PS00225; CRYSTALIN_BETAGAMMA; UNKNOWN_1.
KW Kinase.
SO SEQUENCE 446 AA; 51670 MW; C13A9C6D1EAEE22C CRC64;

Query Match 34.7%; Score 755; DB 2; Length 446;
Best Local Similarity 39.1%; Pred. No. 5.8e-36;
Matches 175; Conservative 88; Mismatches 171; Indels 14; Gaps 5;

```

```

Oy 173 LLIQVLYIESYVNPITLKFKEFYVYIYLLFLLISFSLQYTKQVQNMOKRQAOJR 232
Db 175 VMQSLSYLEBQGIQSTTVHLLVYLLFPMGIKLLTYLKDKLHELNQDRLR 234
Oy 233 ITQYQOIESLYKDIRSRHDYINILTSRLGIENKDLASIEKIYHQILEKTHQLODT 292
Db 235 MERSRHIIEELKYEVSRFDYTNLLTSLRLGIEEDMQIKREIYDSVLKSSSEKLDN 294
Oy 293 RYNIHLANIONDAVKGLSLAKIIEAQNKKIYAVNEVSSKIQLEPEMLDFTILSLCD 352
Db 295 YDLRLVNIHRALKSLACKFKAKDKNIVFVEPEIYQEGVSLDELTVVSLD 354
Oy 353 AIEAFESLNPETQLAFKKNGSIVFIQNSTKEKOIDVSKIFKENYSTKSGNRIGLA 412
Db 355 AIEASVACOPHVAIFKNGAOTETFIENSIRKEGIDISEIFSGASSKGEERGVLTY 414
Oy 413 VNHLEHYPKTSLOTSNHHLLFKOLLII 440
Db 415 VMKIVESHPTSLNTGQDHVFRQVLTV 442

RESULT 3
Oy 315 PRELIMINARY; PRT; 446 AA.
ID 09S115
AC 09S115
DT 01-MAY-2000 (TReMBLrel, 13, Created)
DT 01-MAY-2000 (TReMBLrel, 13, Last sequence update)
DT 01-OCT-2001 (TReMBLrel, 18, Last annotation update)
DE HISTIDINE KINASE (BLPH PROTEIN) (SENSOR HISTIDINE KINASE BLPH,
DE PUTATIVE).
GN HK13 OR IH OR BLPH OR SP0527.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=99453779; PubMed=10524254;
RA Lange R., Wagner C., de Saizieu A.B., Flint N., Molnos J., Stieger M.,
RA Caspers P., Kamber M., Keck W., Amrein K.;
RA "Domain organization and molecular characterization of 13 two-
RT component systems identified by genome sequencing of S. pneumoniae.";
RL Gene 237:223-234(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=KNR7/87;
RA Reichmann P., Hakenbeck R.;
RA "A peptide inducible Signal Transduction System in Streptococcus
RT pneumoniae: Evidence for Bacteriocin Production.";
RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RA MEDLINE=21357209; PubMed=11463916;
RA Tettelein H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., Debey R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwin M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:498-506(2001).

```

DR EMBL; AJ006401; CAB54587.1; -
 DR EMBL; AJ278419; CAC18581.1; -
 DR EMBL; AJ276410; CAC03516.1; -
 DR EMBL; AE007364; AAK74685.1; -
 DR TIGR; SP0527; -
 DR InterPro; IPR001064; Crystalin.
 DR PROSITE; PS00225; CRYSTALLIN_BETACRAMMA; UNKNOWN_1.
 KW Kinase; Complete proteome.
 SQ SEQUENCE 446 AA; 51824 MW; D8CBGD28P5DP274C CRC64;

Query Match 34.5%; Score 749; DB 16; Length 446;
 Best Local Similarity 37.8%; Pred. No. 1.3e-35;
 Matches 170; Conservative 89; Mismatches 173; Indels 18; Gaps 4;

QY 1 MNEAMTL-----SNGLLTYIVFLFLFSSKSNVTL-----KKEFLFSSINFLMIA 51
 DB 1 MNTAMLLYLVNGL-----ETVIFKVDGIDLTFFERIFKFLKLLLAFFVAMIS 52
 QY 52 VTWVNVLPYAPLPYFALSTYLNKNSLSNIFYGLPVSDDLFRRAIFFILDGT- 110
 DB 53 YVGNVYLSYFMEPLVIGLSFLRLGLPKLLEFGLEPMLVNLFYRGVSFVLPFLG 112
 QY 111 OGIVSSITTYMIEFAGIALSYFLSPVNDIGRLDSLRKMKVKRLIPNITMLLY 170
 DB 113 OGQVYDGYSTGICIIIFNFFSLAFKMLDVFSTLREILDKAFQSLQIMMIGCY 172
 QY 171 YLLQLVLYIESYVNIPTLKRKVVVYVLLFLILISFLSQYTRKQVONEIMAKENAI 230
 DB 173 YLVWESLSFEEYQSIOSKTVRHLIIIVFLFFMGVYKLDLYLEKLYERLEQOALRY 232
 QY 231 RNIVYQSOLESYKDIRSPRHDYLNILTSRLGLENKDLASIEKIYHQILEKGHODL 290
 DB 233 RDMRYRRIHELYKKEVSRHDYTNLTLSTLGLFEEDEMDIKEVYSVLKDSQKQIN 292
 QY 291 TRVIGHLANTONDAVKGISAKILEONKKAIVANVSSKIQLEPEMLDPTILSLC 350
 DB 293 NKDGLVLRKALKSLAGKFLKARDKNIVFENVEEIOVGMSLTLFLITVSLIC 352
 QY 351 DNAIEAFESINPEIOLAFPKNGSIIVFTIONSTKEKOIVSKIKENYSTKSGNRIGL 410
 DB 353 DNIEASVEASOPHVSIAFLKNGAQETFIENSKEGIDISEIFSGASSGGERGGL 412
 QY 411 AKVNHILEHYPKTSLOTSNNHHLFKQDLII 440
 DB 413 YTWKIVESHPTSLNTTCOMQVROVLTV 442

RESULT 4
 ID 032832 PRELIMINARY; PRT: 426 AA.
 AC 032832; 01-JAN-1998 (TREMblrel. 05, Created)
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE PUTATIVE HISTIDINE KINASE (EC 2.7.3.-) (ORF8).
 GN PLSK.
 OS Lactobacillus plantarum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
 OC Lactobacillus
 OX NCBI_TaxID=1590;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LPC010;
 RX MEDLINE=98247325; PubMed=9572965;
 RA Stephens S.K., Floriano B., Cathcart D.P., Bayley S.A., Witt V.F.,
 RA Jimenez-Diaz R., Warner P.J., Ruiz-Barba J.;
 RT "Molecular analysis of the locus responsible for production of
 RT plantaricin S, a two-peptide bacteriocin produced by Lactobacillus
 RT plantarum LPC010";
 RL Appl. Environ. Microbiol. 64:1871-1871(1998).
 CC -1 SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 DR EMBL; Y15127; CAA5398.1; -

DR InterPro; IPR003594; HATPase_C.
 DR InterPro; IPR004359; HIS_KIN_sig.
 DR Pfam; PF02518; HATPase_C_1.
 KW Kinase; Transferase; Transmembrane.
 FT TRANSMEM 42 62 POTENTIAL.
 FT TRANSMEM 78 98 POTENTIAL.
 FT TRANSMEM 142 162 POTENTIAL.
 FT TRANSMEM 179 199 POTENTIAL.
 SQ SEQUENCE 426 AA; 48387 MW; 4B203DFC9EC34C0 CRC64;

Query Match 25.1%; Score 545.5; DB 2; Length 426;
 Best Local Similarity 30.3%; Pred. No. 5.7e-74;
 Matches 131; Conservative 99; Mismatches 174; Indels 29; Gaps 7;

QY 18 VLFLFLPSKSNVTLSSKELTFLSINFLMIAVWVNVNLPYAPLPYFALSTIYLNR 77
 DB 11 VTSLLIYHRISSTIKLTMREIALVLFGLTFSSILS---TFPFAVIMMLTAIGIKDGL 66
 QY 78 QMSLSNIFYGLLPVASSDLFRRAIFFILDGTQGIYMS---SIITTYMIEFAGIAL 132
 DB 67 HQKYLAFPPYSY-----SVTFEFLVGNLSLGFVGDENIYIPFTALPVI 116
 QY 133 SYLFSEVNVNIGRLKDSLRKMKVKRLIPNITMLLYLQVLYVSSYVNIPTL--- 189
 DB 117 NELICKLNPDPDFLRNV-GATRNLYGVN--FIFSLCTVQY--GSYTWEDTMGDI 170
 QY 190 -KRRFVYVYVLLFLILISFLSQYTRKQVONEIMAKENAIINTQYSOQIESLYKDIR 248
 DB 171 GYVRVYLGSLFLLIALVILVILNKLASIOOSQRCREEQIHMQRYSQIENLTVEVR 230
 QY 249 SPRHDYLNILTSRLGLENKDLASIEKIYHQILEKTHODRNVNIGHLANIONDAVG 308
 DB 231 SRHDYINMLSLSGIDENITEKTYNVNLSGLNASIQSSKYSILNSKLVPAIS 290
 QY 309 IISAKILEONKKAIVANVSSKIQLEPEMLDPTILSLCDNAIDAFESINPEIQIA 368
 DB 291 IISKRIYAQONGIEVNISSDVTYTFDLDIRVIAIFLNAIEASLTDLPMSVA 350
 QY 369 FPKNGSIIVFTIONSTKEKOIVSKIKENYSTKSGNRIGLAKVNHILEHYPKTSLOTS 428
 DB 351 FIEEDDSQVILVYNNAPSYIDKRRIPESGRTGHNHRIGLATVMDLKNKYNASLEIE 410
 QY 429 NNHHLFKQDLII 441
 DB 411 FRNSLFTOKLAIK 423

RESULT 5
 ID 09A1J7 PRELIMINARY; PRT: 448 AA.
 AC 09A1J7; 01-JUN-2001 (TREMblrel. 17, Created)
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE PUTATIVE HISTIDINE KINASE, POSSIBLY INVOLVED IN COMPETENCE.
 GN SPV0242.
 OS Streptococcus pyogenes.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus
 OX NCBI_TaxID=1314;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SF370 / ATCC 700294 / SEROTYPE M1;
 RX MEDLINE=21192684; PubMed=11296296;
 RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
 RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
 RA Qian Y., Jia H.G., Najjar F.Z., Ren O., Zhu H., Song L., White J.,
 RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
 RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
 DR EMBL; AE006491; AAK33320.1; -
 DR InterPro; IPR003594; HATPase_C.

DR InterPro: IPR003860; Phosphopantl_attach.
 DR Pfam: PF02518; HATPase_C_1.
 DR SMART: SM00387; HATPase_C_1.
 DR PROSITE: PS0012; PHOSPHOPANTHEINE; UNKNOWN_1.
 KW Kinase; Complete proteome.
 SQ SEQUENCE 448 AA; 52221 MW; 230C003A104CBF5 CRC64;

Query Match 24.8%; Score 539.5; DB 16; Length 448;
 Best Local Similarity 29.98; Pred. No. 1.3e-23;
 Matches 137; Conservative 105; Mismatches 173; Indels 43; Gaps 11;

QY 2 NEALMILNSGLTYLTVFLLEFSKSVNTLSKRELTFESINFLIMIAVTWVNL 60
 Db 4 SEYMFLLSLIYYMT---KIYFSLSDITLTPWKQTLIALALFF-----NCF 50
 QY 61 ----YPAELXFLALSTYLNKNSLNIYFGLLPVASSDLFR---AIFFLIDGTG 112
 Db 51 PYLSPLLIDPLFLVLRQETKQLFSLKALFLVAPSVLVDLSRFMGCTIVIPYFLSG 110
 QY 113 IVMGSSIT---ITTYMIEFAGIALSYLFLSVFNVDIGRLKDSLTKMKYKRLIPNITMLY 170
 Db 111 IYGLIFFLDYLAVLILFESFA-----IINYMGKDKYMKCSGSKRSHNFYCTLMF 163
 QY 171 YLLIYV-LVYIESYVPIPLKFR-----KEVVIYVYLFLFLILISFLOTKKON 220
 Db 164 VLVIYVDFVILGF-TDPLLFHSHSLVYPTPYKILFLMFLILLVILSTFNHSSKEYLYN 222
 QY 221 EIMAKQKQIRNITQYISOIESLYKDIRSPRDYLNITLSRLGIENKDLASTIKYHOI 280
 Db 223 ELREBOAAVMTLEIYGHLEKLYRDVAFOSDYLSRIERLGOAISESTIQDIYAO 282
 QY 281 LKRTGQLQDTRYNIGHLANIONDAVKGILSAKILEAONKKNVYSSKIQLOPME 340
 Db 283 VHEANDYWDKRYNISKRLKINISSIKLSAKIISKEKSIDLNVVDPNIKETIPEL 342
 QY 341 DFTILSLICDVAIEAFESLNPETQLAEFRKNGSIVFIIONSTKEKQIDVSKIRENS 400
 Db 343 DLLMSIFCDVAIEAFESLNPETQLAEFRKNGSIVFIIONSTKEKQIDVSKIRENS 401
 QY 401 TRGNSRGIGLAKVNHLEHYPKTSLOTNSHHHLFKOLL 438
 Db 402 SKSGERGIGLSMAQRILKKPYLSIRTSFPEKRSQTL 439

RESULT 6
 Q99YP4 PRELIMINARY; PRT; 451 AA.
 AC Q99YP4; 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE PUTATIVE TWO-COMPONENT RESPONSIBLE HISTIDINE KINASE.
 GN SPY1005.
 OS Streptococcus pyogenes.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 NCBI_TaxID=1314;
 RN [1]
 RP SEQUENCE FROM N. A.
 RC STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;
 MEDLINE=21192684; PubMed=11296296;
 RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
 RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
 RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
 RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
 RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
 DR EMBL: AE006592; AAK34382.1;
 KW Kinase; Complete proteome.
 SQ SEQUENCE 451 AA; 52415 MW; 045D080E66D45B97 CRC64;

Query Match 24.2%; Score 526; DB 16; Length 451;
 Best Local Similarity 30.08; Pred. No. 8e-23;
 Matches 136; Conservative 86; Mismatches 211; Indels 20; Gaps 5;

QY 1 NEALMILNSGLTYLTVFLLEFSKSVNTLSKRELTFESINFLIMIAVTWVNL 60
 Db 2 LEDPFLQFLGFIPLDIIEMLTLKFSFVSAIPRLKNIPLSLSMVLPQVFW-----AF 56
 QY 61 YP-----AEPYFLALSTYLNKNSLNIYFGLLPVASSDLFRALFTI--LD 108
 Db 57 FPDHFLDVMLAOPFLPALALYKGSIRAKFLMFAFFPLVISLVRFVFMPLF 116
 QY 109 GQGIYMGSSITTYMIEFAGIALSYLFLSVFNVDIGRLKDSLTKMKYKRLIPNITML 168
 Db 117 GMYSVVKNHNTLIYSICFSLPIYRCIOYFHFDFSTWQYFQSHRASKLVLVTNSSMA 176
 QY 169 LYLLIYQVLYE-SYVNIPLKFRFVYVYLIPILISFLOTKKONELMAOKE 227
 Db 177 LYLLIYQVLYE-SYVNIPLKFRFVYVYLIPILISFLOTKKONELMAOKE 236
 QY 228 AQIRNITQYISOIESLYKDIRSPRDYLNITLSRLGIENKDLASTIKYHOI 287
 Db 237 --YRELINSQHLGLYODIOLRLLTYSRRKIGIEQNDISIVRLYEGILMAEKN 294
 QY 288 LQDTRYNIGHLANIONDAVKGILSAKILEAONKKNVYSSKIQLOPME 347
 Db 295 AKQDRIDLCTCDLQVEIRHIVLAKLEAKNKLAKEVSIKPCITVTFLEVDFTKLUS 354
 QY 348 ILCDNAIEAFESLNPETQLAEFRKNGSIVFIIONSTKEKQIDVSKIRENS 407
 Db 355 FLIDNAIEMSLTEPKPCISIFLQNHVIVIOSSTKQDOSQVFAIPALKRRDMQ 414
 QY 408 IGLAVNHLEHYPKTSLOTNSHHHLFKOLL 440
 Db 415 FDLRNVTTILNRYDYLTISQIHGILTLQLEI 447

RESULT 7
 Q9A1J6 PRELIMINARY; PRT; 427 AA.
 AC Q9A1J6; 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE PUTATIVE HISTIDINE KINASE.
 GN SPY0244.
 OS Streptococcus pyogenes.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 NCBI_TaxID=1314;
 RN [1]
 RP SEQUENCE FROM N. A.
 RC STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;
 MEDLINE=21192684; PubMed=11296296;
 RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
 RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
 RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
 RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
 RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
 DR EMBL: AE006491; AAK33321.1;
 KW Kinase; Complete proteome.
 SQ SEQUENCE 427 AA; 49908 MW; 01C9D236EAD7A0D6 CRC64;

Query Match 23.8%; Score 517; DB 16; Length 427;
 Best Local Similarity 27.6%; Pred. No. 2.5e-22;
 Matches 122; Conservative 110; Mismatches 184; Indels 26; Gaps 6;
 QY 5 LMLNSGLTYLTVFLLEFSKSVNTLSKRELTFESINFLIMIAVTWVNLFP- 62
 Db 1 MMYIYSMALEFATPIINSVIFARVASIKLSWKVSI-----IGISFYIANKIDK 51

QY 63 ---AEPYIAIATSLYNQNSLSLNIFFYGLPVASSDLFRRAIFFIIDGTGIVWSS1 119
 DB 52 VILIDQLEFFIVSLISAPKKLFEHMFNGFFITLIVELFVIGSIFPLPAVLGSIQIN 111
 QY 120 ITTYMIEFA---GIALSYLSEVFNVDIGRLKDSITKAKVKRLIPNNITMLLYLLIQV 176
 DB 112 NNKLELCYLPVLPFVLFYFISFIDLSLR-FISECKMKKVMFMMNTAMFSYFPAHF 170
 QY 177 LVIESYNTPLKFRKREYIYLLFLILISLSQITKQVQNEIQAQIRNITQY 236
 DB 171 LVYQSGFALYQYRSLVFLYLAIFTWIVYKIDREFKQDLSQKLTQAOERLAVLENY 230
 QY 237 SQOIEYLDISFRHDYINILITSLRLGIENKDLASIEKIYHQILEKTHOLOPTRNIG 296
 DB 231 NQSIQYLEIRTVKHDSENLISLKDSDIGDIDLIRVYDVYIQQSATSMNTNYS 290
 QY 297 HIANIONDAVKILSAKILEAONKRIAVNESSKIQLEPEMELDTITLISLCDNAIEA 356
 DB 291 SLDNIEAVIRSMNSKLEAOYIGILEYIEIPVDIHLPLKIDLVLEFTGLVDNIET 350
 QY 357 AFESLNEIQLAFKNGSIVFIIONSTRKQIDVSKIFENYSTKSGNKGIGLAKVNH1 416
 DB 351 AKGSRPEFISIVFKQDNKQLEFIENSTKTRKVDIAKRFDAQOQNS-----AHLIV 402
 QY 417 LEHYPKTSQTSNHHLEKOLL 438
 DB 403 LQSYQITVLTSDHYRLKOLL 424

RESULT 8
 Q97MM1 PRELIMINARY: PRT: 447 AA.
 ID AC Q97MM1: 01-OCT-2001 (TREMBlrel. 18, Created)
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE HISTIDINE KINASE-LIKE APPASE;
 GN CAC0080.
 OS Clostridium acetobutylicum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OX NCBI_TaxID=1488;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 824 / DSM 792 / VYM B-1787;
 RX MEDLINE=21359325; PubMed=11466286;
 RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
 Gibson R., Lee H.M., Dubois J., Olu D., Hitti J., Wolf Y.I.,
 RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
 RA Bennett G.N., Koonin E.V., Smith D.R.,
 RT "Genome sequence and comparative analysis of the solvent-producing
 bacterium Clostridium acetobutylicum."
 RL J. Bacteriol. 183:4823-4838(2001).
 DR EMBL: AE007520; AAK78065.1;
 DR InterPro: IPR003594; HATPase_C.
 DR SMART: SMO0387; HATPase_C.1.
 KW Kinase; Complete proteome.
 SQ SEQUENCE 447 AA; 51289 MW; A0352A687A692F7B CRC64;

Query Match 20.9%; Score 455; DB 16; Length 447;
 Best Local Similarity 27.4%; Pred. No. 9.4e-19;
 Matches 124; Conservative 96; Mismatches 163; Indels 70; Gaps 10;

QY 11 GLLTYLVFLFLFESKVSNTLSKKELETSISNFLIMIAVTWNVNLFYPAPEFIA 70
 DB 33 GMVITEIISLFLM-----LGMQOYVMA-----YLI 61
 QY 71 LSTYLRKNS---LSLNTYGLPVASS-----DLFRAIFFIIDGTGIVWSSIT- 120
 DB 62 PCIFMFLQTSVDVLIS-----LPVAGCLIIIVMDYFLSVYVNVFVNVSVIKNDSIH 116
 QY 121 -TYMIEFAGIALSYLSEVFNVDIGRLKDSITKAKVKRLIPNNITMLLYLLIQVLYV 179

DB 117 WITTEIEFIVYSTRIRKREVD-----KRAYNGHKGITGVAAAGTLLIMFEV 167
 QY 180 IESYVNI-----PTLKRREYIV-YLLIFLILISLSQYKQKQVQNEIQAQIRNIT 228
 DB 168 FYNNVAVAPNNSVNSNTWRKIGIVLFESYAILLILIVIRTLIRGIRKME---LKSNE 224
 QY 229 QIRNITQYSQOIESLYNDISFRHDYINILITSLRLGIENKDLASIEKIYHQILEKTHO 288
 DB 225 EFQSLQETYNKLEKLDKMGFRHDYINILISLMAGYIONRDEGLEFFDDKIMPLSKAM 284
 QY 289 QDTRNIGHLANIONDAVKILSAKILEAONKRIAVNESSKIQLEPEMELDTITLISI 348
 DB 285 KSNNEKIGLQINIEVPEIKGWFSAKIIIRAQETGIDYIDVAESIKSFNEIIDLRSVIGI 344
 QY 349 LCDNAIEAFESLNEIQLAFKNGSIVFIIONSTRKQIDVSKIFENYSTKSGNKGIGL 408
 DB 345 LLDNAIEASEKCDRPSMKVAIVNKRKSVMTVILNNNEIPIYKIKRGSTGDKDNKGI 404
 QY 409 GLAKVNHILEHYPKTSQTSNHHLEKOLLIK 441
 DB 405 GLSNLKDIIGKYPVMDLVYEDNQFKQIDIK 437

RESULT 9
 P95767 PRELIMINARY: PRT: 453 AA.
 ID AC P95767: 01-MAY-1997 (TREMBlrel. 03, Created)
 DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE HISTIDINE PROTEIN KINASE.
 GN COMD.
 OS Streptococcus gordonii challis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OX Streptococcus.
 OX NCBI_TaxID=29390;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CHALLIS;
 RX MEDLINE=97294448; PubMed=9150204;
 RA Lunford R.D., Robie A.G.,
 RT "comY, a gene similar to comGA of Bacillus subtilis, is essential for
 competence-factor-dependent DNA transformation in Streptococcus
 RT gordonii."
 RL J. Bacteriol. 179:3122-3126(1997).
 DR EMBL: U80077; AAC45306.1; -.
 KW Kinase.
 SQ SEQUENCE 453 AA; 53636 MW; 781D5360489B0139 CRC64;

Query Match 20.9%; Score 455; DB 2; Length 453;
 Best Local Similarity 28.9%; Pred. No. 9.5e-19;
 Matches 131; Conservative 94; Mismatches 179; Indels 50; Gaps 12;

QY 1 MNEALMISNGLTYLVFLFESKVSNTLSKKELETSISNFLIMIAVTWNVNLFYPAPEFIA 59
 DB 32 INIKILIFKWLINWISLTSFL-----LLPDWKFRFRIHFYHLL----- 74
 QY 60 FYPAPEFIALSYLNRKNSLSLNTYGL---LPVASSDLFRRAIFFIIDGTGIVWSSIT 116
 DB 75 ---IOPLFYRYFFPKRKREYENYLNFLSEVYLSVETSEFSLVIT-----SSITG 123
 QY 117 SSIYTT-YMIEFAGIALSYLSEVFNVDIGRLKDSITKAKVKRLIPNNITMLLYLLIQV 174
 DB 124 DSFVQOHNDIFIIYIINLSLFLILKVDPDFEFYFYKKEPIYNDLYNNKSGYIVIHILL 183
 QY 175 QVLY-----VIESY-NVITPLKFRKREYIYLLFLILISLSQITKQVQNEIQAQIRNIT 227
 DB 184 NISHFSENAHNSFASATATIGF-----IMFSLTFYKASAREYKAK---EIQQKE 235
 QY 228 AQIRNITQYSQOIESLYNDISFRHDYINILITSLRLGIENKDLASIEKIYHQILEKTHO 287

Db 236 EQ-ROLYTDEIVGVLNIEIRGRHDAAGMLTSLQTSINGSDMKKEVERIFHNVLQSANIS 294

QY 288 LQDTRYNIGHLANIONDAVKIGILSAKLEIAONKRIAVNEVSSKIOLPEMELIDFTIIS 347

Db 295 LRSDDYTFEFLNNVODTALRSVLQITIFKARECGVEIYFEKKDVIELPKMLDLVAVAS 354

QY 348 ILCDNAIEAFESLAPETIOLAFPKKNGSIYFIIONSTREKCIDVSKTFEKENSTKGSNG 407

Db 355 VILNNVAGAAESPCKTMNVSLVLDKEIYVIONSRQSRINLEIYEYGFSTGKENG 414

QY 408 IGLAKVNHILEHYPKTSLOTSNHHHLFKOLLITK 441

Db 415 LGLNNVKEIIDKYDEVILETIDETNYFIOVVRK 448

RESULT 10

P72446 PRELIMINARY; PRT; 453 AA.

AC P72446: PRELIMINARY; PRT; 453 AA.

DT 01-FEB-1997 (TREMblrel. 02, Created)

DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)

DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)

DE HISTIDINE KINASE.

GN COMD1.

OS Streptococcus gordonii.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Streptococcus.

OX NCBI_TaxID=1302;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CHALLIS (NCTC 7868);

RX MEDLINE=97032151; PubMed=8878047;

RA Haverstein L.S., Gaustad P., Nes I.F., Morrison D.A.;

RT "Identification of the streptococcal competence-pheromone receptor.";

RL Mol. Microbiol. 21:863-869(1996).

DR EMBL: X98109; CAA66787.1; -.

KW kinase.

SQ SEQUENCE 453 AA; 53633 MW; 281D55CDE4063E2 CRC64;

Query Match 20.7%; Score 450; DB 2; Length 453;

Best Local Similarity 28.6%; Pred. No. 1.8e-18; Mismatches 180; Indels 50; Gaps 12;

Matches 130; Conservative 94; Mismatches 180; Indels 50; Gaps 12;

QY 1 MNEALMISNGILTYLVFLFELFSKYSNVLTKKELTLF-SISNPLIMATMVNVL 59

Db 32 INKILILFWGLIMVISELYSLF-----LDPYWKIFRDIYHLYFL----- 74

QY 60 FVPAPLYIALSIYLNONSLSNIFYGL--LPVASSDLFRRAIFFILDTGQIVMG 116

Db 75 ----IQPLFYKFLIKREYENYVNLFSYIYLSVETSTFSLVIL-----SSITG 123

QY 117 SSIITF-TWIEFAGIALSYFLSVFNVDIGRLKDSLTKMKY-KRLIPMNTMLLYLTL 174

Db 124 DSVFQKHVDIFILINLSLFIILKVDFEYFEYKKEPYKNDLVNMSYVLIHL 183

QY 175 QVLY-----VIESY-NVIFPLKFRKRVYIYLLFLLISLSQYTKOKVONIMAKE 227

Db 184 NISHMSENAHLNFSAMATIGF-----IMFXTLFLYKARSQRYEKAK--EIQCKE 235

QY 228 AQIRNITQSOQIESLYKDIRFRDYNIILTSRLGIENKDLASIEKIYHOLEKTHQ 267

Db 236 EQ-ROLYTDEIVGVLNIEIRGRHDAAGMLTSLQTSINGSDMKKEVERIFHNVLQSANIS 294

QY 288 LQDTRYNIGHLANIONDAVKIGILSAKLEIAONKRIAVNEVSSKIOLPEMELIDFTIIS 347

Db 295 LRSDDYTFEFLNNVODTALRSVLQITIFKARECGVEIYFEKKDVIELPKMLDLVAVAS 354

QY 348 ILCDNAIEAFESLAPETIOLAFPKKNGSIYFIIONSTREKCIDVSKTFEKENSTKGSNG 407

Db 355 VILNNVAGAAESPCKTMNVSLVLDKEIYVIONSRQSRINLEIYEYGFSTGKENG 414

QY 408 IGLAKVNHILEHYPKTSLOTSNHHHLFKOLLITK 441

Db 415 LGLNNVKEIIDKYDEVILETIDETNYFIOVVRK 448

RESULT 11

O33646 PRELIMINARY; PRT; 453 AA.

AC O33646: PRELIMINARY; PRT; 453 AA.

DT 01-JAN-1998 (TREMblrel. 05, Created)

DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)

DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)

DE HISTIDINE KINASE.

GN COMD.

OS Streptococcus gordonii.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Streptococcus.

OX NCBI_TaxID=1302;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NCTC 3165;

RX MEDLINE=96012953; PubMed=9352904;

RA Haverstein L.S., Hakenbeck R., Gaustad P.;

RT "Natural competence in the genus Streptococcus: Evidence that streptococci can change phenotype by interspecies recombinational exchanges.";

RL J. Bacteriol. 179:6589-6594(1997).

DR EMBL: AJ000870; CAA04357.1; -.

KW kinase.

SQ SEQUENCE 453 AA; 52756 MW; ABF273293A2CC52 CRC64;

Query Match 20.7%; Score 449.5; DB 2; Length 453;

Best Local Similarity 28.2%; Pred. No. 2e-18; Mismatches 173; Indels 59; Gaps 10;

Matches 127; Conservative 91; Mismatches 173; Indels 59; Gaps 10;

QY 12 LITVITVFLFELFSKYSNVLTKKELTSLISNPL-IMATMVNVLFPAPLYRI 69

Db 38 IFVLTIFSSLLITQFYNLIPYGDV----LSFTLYLTLIFQPLILHYSKKGILYG 93

QY 70 ALSTYLNONSLSNIFYGLLPVASSDLFRRAIFFILDTGQIVMGSSITTY----- 123

Db 94 YVSIF-----LSLLIY---LSVSSSETFSSVIL-----SSVNGDFVNOVMSYYT 136

QY 124 -----MTEFAGIALSYFLSVFNVDIGRLKDSLTKMKYKRLIPMNTMLLY 171

Db 137 IVNIALFVYKSEFEFEFSYKNDFEKEI-----MNVKIYVTHIILINISH 187

QY 172 LLIQVLYIESYNYIPLKFRKRVYIYLLFLLISLSQYTKOKVONIMAKEAQR 231

Db 188 WFSNAHLNFSAMATIGF-----IMFSLTFYKARSQRYEKAK--EIQCKEEO-R 238

QY 222 NITQVSOQIESLYKDIRFRDYNIILTSRLGIENKDLASIEKIYHOLEKTHQLODT 291

Db 236 EQ-ROLYTDEIVGVLNIEIRGRHDAAGMLTSLQTSINGSDMKKEVERIFHNVLQSANIS 294

QY 292 RYNIGHLANIONDAVKIGILSAKLEIAONKRIAVNEVSSKIOLPEMELIDFTIIS 347

Db 295 LRSDDYTFEFLNNVODTALRSVLQITIFKARECGVEIYFEKKDVIELPKMLDLVAVAS 354

QY 352 NATEAFESLAPETIOLAFPKKNGSIYFIIONSTREKCIDVSKTFEKENSTKGSNG 407

Db 355 VILNNVAGAAESPCKTMNVSLVLDKEIYVIONSRQSRINLEIYEYGFSTGKENG 414

QY 412 KVNHILEHYPKTSLOTSNHHHLFKOLLITK 441

Db 419 NVKEIIDKYDEVILETIDETNYFIOVVRK 448

RESULT 12

O33597 PRELIMINARY; PRT; 448 AA.

AC O33597: PRELIMINARY; PRT; 448 AA.

DT 01-JAN-1998 (TREMblrel. 05, Created)

RESULT	13		
O9X995			
ID	O9X995;		PRT; 441 AA.
AC	O9X995;		
DT	01-NOV-1999 (TREMBLrel_12,		
DT	01-NOV-1999 (TREMBLrel_12,		
DT	01-DEC-2001 (TREMBLrel_19,		
DE	COND. PROTEIN.		
GN	COND.		
OC	Streptococcus oralis.		
Bacterla:	Firmicutes; Bacillus/Clostridium group; Streptococcaceae;		

RESULT	14	
09X990		
ID	09X990	PRELIMINARY; PRT; 441 AA.
AC	09X990;	
DT	01-NOV-1999 (TREMBLrel. 12, Created)	
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)	
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)	
DE	COMM. PROTEIN.	
GN	COMM.	
OS	Streptococcus mitis.	
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;	
OC	Streptococcus	
OX	NCBI_TaxID=28037;	
RN	[1]	

